

Wed Oct 10 07:44:15 2001

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:08:05 ; Search time 1666.31 Seconds  
(without alignments)  
232.066 Million cell updates/sec

Title: US-09-396-196F-1  
Perfect score: 25  
Sequence: 1 gacatgtcgcaagtcacagaatta 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_ba3.\*
- 4: gb\_in1.\*
- 5: gb\_in2.\*
- 6: gb\_in3.\*
- 7: gb\_om.\*
- 8: gb\_ov.\*
- 9: gb\_pat1.\*
- 10: gb\_pat2.\*
- 11: gb\_ph.\*
- 12: gb\_pl1.\*
- 13: gb\_pl2.\*
- 14: gb\_pl3.\*
- 15: gb\_pl4.\*
- 16: em\_ba1.\*
- 17: em\_ba2.\*
- 18: em\_fun.\*
- 19: em\_htgo\_hum.\*
- 20: em\_htgo\_inv.\*
- 21: em\_htgo\_rod.\*
- 22: em\_htg\_hum1.\*
- 23: em\_htg\_hum2.\*
- 24: em\_htg\_hum3.\*
- 25: em\_htg\_hum4.\*
- 26: em\_htg\_hum5.\*
- 27: em\_htg\_hum6.\*
- 28: em\_htg\_hum7.\*
- 29: em\_htg\_hum8.\*
- 30: em\_htg\_inv1.\*
- 31: em\_htg\_inv2.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_rod.\*
- 34: em\_hum1.\*
- 35: em\_hum2.\*
- 36: em\_hum3.\*
- 37: em\_hum4.\*
- 38: em\_hum5.\*
- 39: em\_hum6.\*
- 40: em\_hum7.\*
- 41: em\_in.\*
- 42: em\_om.\*
- 43: em\_or.\*

- 44: em\_ov.\*
- 45: em\_pat.\*
- 46: em\_ph.\*
- 47: em\_pl.\*
- 48: em\_ro.\*
- 49: em\_sts.\*
- 50: em\_sy.\*
- 51: em\_un.\*
- 52: em\_vl.\*
- 53: gb\_sts1.\*
- 54: gb\_sts2.\*
- 55: gb\_sts3.\*
- 56: gb\_sy.\*
- 57: gb\_un.\*
- 58: gb\_vil.\*
- 59: gb\_vil2.\*
- 60: gb\_htg1.\*
- 61: gb\_htg2.\*
- 62: gb\_htg3.\*
- 63: gb\_htg4.\*
- 64: gb\_htg5.\*
- 65: gb\_htg6.\*
- 66: gb\_htg7.\*
- 67: gb\_htg8.\*
- 68: gb\_htg9.\*
- 69: gb\_htg10.\*
- 70: gb\_htg11.\*
- 71: gb\_htg12.\*
- 72: gb\_htg13.\*
- 73: gb\_htg14.\*
- 74: gb\_htg15.\*
- 75: gb\_htg16.\*
- 76: gb\_htg17.\*
- 77: gb\_htg18.\*
- 78: gb\_htg19.\*
- 79: gb\_htg20.\*
- 80: gb\_htg21.\*
- 81: gb\_htg22.\*
- 82: gb\_htg23.\*
- 83: gb\_htg24.\*
- 84: gb\_htg25.\*
- 85: gb\_pr1.\*
- 86: gb\_pr2.\*
- 87: gb\_pr3.\*
- 88: gb\_pr4.\*
- 89: gb\_pr5.\*
- 90: gb\_pr6.\*
- 91: gb\_pr7.\*
- 92: gb\_pr8.\*
- 93: gb\_pr9.\*
- 94: gb\_rol.\*
- 95: gb\_rol2.\*
- 96: gb\_in4.\*
- 97: gb\_pr10.\*
- 98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	25	100.0	1041	9 AR029499	Sequence
2	25	100.0	1041	9 AR034916	Sequence
3	25	100.0	1084	9 A11530	Biob gene o
4	25	100.0	1121	10 E00893	Genomic DNA
5	25	100.0	5793	2 E00893	E.coli 7,8-
6	25	100.0	5872	9 A38246	Sequence 1
7	25	100.0	5872	9 A38251	Sequence 6
8	25	100.0	5872	9 A93674	Sequence 1

## RESULT 2



Db	44	GACATTGTCGCAAGTCACAGAATTA	68
RESULT	4		
LOCUS	E00893	1121 bp	DNA
DEFINITION	Genomic DNA encoding biotin Synthetase.	PAT	29-SEP-1997
ACCESSION	E00893		
VERSION	E00893.1	GI:2169154	
KEYWORDS	JP 1986149091-A/1.		
SOURCE	Escherichia coli.		
ORGANISM	Escherichia coli		
	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
	Escherichia.		
REFERENCE	1 (bases 1 to 1121)		
AUTHORS	Hirono,Y., Kojima,T. and Kimura,H.		
TITLE	DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND PRODUCTION OF BIOTIN		
JOURNAL	Patent: JP 1986149091-A 1 07-JUL-1986;		
	NIPPON SODA CO LTD		
COMMENT	OS Escherichia coli		
	PN JP 1986149091-A/1		
	PD 07-JUL-1986		
	PF 24-DEC-1984 JP 1984272605		
	PI HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC		
	C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC		
	C12R1:19);		
	CC strandedness: Double;		
	CC topology: Linear;		
	CC hypothetical: No;		
	CC anti-sense: No;		
	CC *source: strain=Escherichia coli Nsl01;		
	CC Feature is identified by experimental;		
	FH Key		
	FH Location/Qualifiers		
	FT CDS	42..1079	
	FT /product=biotin synthetase'.		
FEATURES	Location/Qualifiers		
source	1..1121		
	/organism="Escherichia coli"		
	/db_xref="taxon:562"		
BASE COUNT	289 a 296 c 325 g 211 t		
ORIGIN			
Query Match	100.0%; Score 25; DB 10; Length 1121;		
Best Local Similarity	100.0%; Pred. No. 0.043;		
Matches	25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 gacattgtcgcaagtcacagaatt 25		
Db	62	GACATTGTCGCAAGTCACAGAATTA	86
RESULT	5		
LOCUS	ECOBIO	5793 bp	DNA
DEFINITION	E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC protein, and dethiobiotin synthetase (bioD), complete cds.		
ACCESSION	J04423		
VERSION	J04423.1	GI:145422	
KEYWORDS	7,8-diamino-pelargonic acid aminotransferase; 7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene; bioC gene; bioD gene; bioF gene; bioG gene; biotin synthetase; dethiobiotin synthetase.		
SOURCE	Escherichia coli (strain K-12) DNA.		
ORGANISM	Escherichia coli		
	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
	Escherichia.		
REFERENCE	1 (bases 1 to 5793)		
AUTHORS	Otsuka,A.J., Buoncristiani,M.R., Howard,P.K., Flamm,J. and		

```

/codon_start=1
/transl_table=11
/protein_id="AAA23517.1"
/db_xref="GI:145427"
/translation="MATVKNQIAAFAAGRAAAHYEQHADLQKRSADALIAMLPRKYT
HVLDAACGPGMSRHRERHAQVTDLDLSPMLVQARQKDAADHYLAGDIESLPLATA
TFDLANSVLAVQWCGNLSTALRELYRVVRPKGVAFTTLVQGLSPERHQACVADPER
HANRFLPPDEIEQSLNGVHYQHIOPIITLWFDLDSAMRSLKGIGATHLHGRDPRIL
TRSQLRLQALWPPQGGRYPLTYHLFLGVARE"
4938..5597
/gene="biob"
/CDS
4938..5597
/gene="biob"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase"
/protein_id="AAA23518.1"
/db_xref="GI:145428"
/translation="MSKRYFVTGDTVEGKTIVASCALLOAKRAGYRTAGYKPVASGS
EKTPELRNSDALQIRNSSQLDYATVNPYTFABPTSPHIIISAGEGRPISLVMSAG
LRALQOQADWLVEGAGGWFTPLSDTTFADWVTQEQLPVLVVGKLGICINHAMLTA
QVQHAGLTLAGVANDVTPPGKRHAETMTILTRMIPRCHWERSFGLQIKQMRQPES
T"
BASE COUNT 1363 a 1554 c 1631 g 1245 t
ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match 100.0%; Score 25; DB 2; Length 5793;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaatta 25
|||||
Db 2032 GACATTGTCGAAGTCACAGAATTA 2056

RESULT 6
LOCUS A38246 5872 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 1 from Patent WO9408023.
ACCESSION A38246
VERSION A38246.1 GI:2294844
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 5872)
Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
Patent: WO 9408023-A 1 14-APR-1994;
LONZA AG (CH)
Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 9500809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
FEATURES
source Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A-15/9"
1..96
/function="PROMOTER PTAC"
/evidence=experimental
23..28
-35_signal
-10_signal 45..49
/standard_name="PROMOTER PTAC"
/evidence=experimental

```

---

```

RBS
105..119
/standard_name="BIOB RBS NO.9"
/evidence=experimental
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"
/codon_start=1
/transl_table=11
/number=1
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAA02325.1"
/db_xref="GI:2294845"
/translation="MAHRPRWTLSQVTELFKPLDLLFEAQOVHRHQHFDPRQVOYST
LLSIKTCACPEDCKYCQSSRYKTKGLEAEERLEMEQVLESARKAKAGSTRFCMGAAW
KNPHERMPYLEOMVOGVKAMGLEACMTLGTLSAQRLANAGLLYYNHNLDTSPEF
YGNLITRTYQERLDILEKVRDAGIKVCSGGIVGIGETVKDRAGLLQLLANLPTPES
VPINMLVKYGTPLADNDVDVDFRTIATVARIMMPTSVRLSAGREOMNQOTQAMC
FMAGSIFYGCKLLTTPNPEDKDLQFLFKLGNLPQQTAVLAGDNEQOORLEQALMT
PDTDEYNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/number=3
/product="PROTEIN"
/protein_id="CAA02326.1"
/db_xref="GI:2294846"
/translation="MATVKNQIAAFAAGRAAAHYEQHADLQKRSADALIAMLPRKYT
HVLDAACGPGMSRHRERHAQVTDLDLSPMLVQARQKDAADHYLAGDIESLPLATA
TFDLANSVLAVQWCGNLSTALRELYRVVRPKGVAFTTLVQGLSPERHQACVADPER
HANRFLPPDEIEQSLNGVHYQHIOPIITLWFDLDSAMRSLKGIGATHLHGRDPRIL
TRSQLRLQALWPPQGGRYPLTYHLFLGVARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:
8-AMINO-7-OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/number=5
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAA02327.1"
/db_xref="GI:2294847"
/translation="MTTDDLAFDQRHIWHPYTSMTSPLVPYVVSABGCELIISDGRR
LVDGSSWMAAIGHYNHQPQNAKMSQIDAMSHVMFGGITHAPAIELCRKLVAMTPOP
LECVELADSGSVAVEMKMALOYWAKGEARQRLFTRNHYGDTFTGMSVCDPDNS
MHLWKGYLPENLFAPAQOSMDGENDERMVGFARLMAAHRHIEAIVILEPVOGAG
GMVMVHPWLKRIKICDREGILLIADIEATGEGRTCKLFACAEHAETAPDILCLGKAL
TGGTMTLSATLTREVAETISNGEAGCFMGHPTFMGNPLACAAANASLILESQDWOO
QVADIEVQLRQALPARKDAENKVRVIGAGVGVVETTHPVNMAALQKFFVEQGVWIRP
FGKLIYLMPPYIILPQQLRLTAANRAVQDETFFCQ"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/number=6

```

```
/evidence=experimental
/product="PROTEIN"
/protein_id="CAA02328.1"
/db_xref="GI:2294848"
/translation="MKLISNDRDGDKLPHRHVNGMGYDGDNI SPLHNDYDPAGTK
SFVTCYDPDAPTSQGMHVVVNLPAADTVLPQFGSLVAMPDGVLTQRTDFGKYG
YDGAAPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFSLASATAMFS"
stem_loop 5583..5605
terminator 5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgctcgaagtcacagaatta 25
|||||
Db 137 GACATTGTCGCAAGTCACAGAATTA 161

RESULT 7
A38251 5872 bp DNA PAT 05-MAR-1997
LOCUS
DEFINITION Sequence 6 from Patent WO9408023.
ACCESSION A38251
VERSION A38251.1 GI:2294849
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia
1 (bases 1 to 5872)
Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
Patent: WO 9408023-A 6 14-APR-1994;
LONZA AG (CH)
Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 950809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
Other publication Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAA02329.1"
/db_xref="GI:2294850"
/translation="MSWQEKINAAIDARRADALRRYPVAQAGRWLVADDRQYLNLF
SSNDYLGUSHHPQIRAWQQAQGFGLGSGSGHVSQYVYVHQAEELEAENLQYSEA
LFTSGFAANQAVTAAAKEDRIAADRSHASLLEAAKLSPSQLRRFAHNDVHLAR
LLASPCQGMVVTVEGFSMDGDSAPLAEITQQVTHQHWLMDVDDAHGTGVIGQGRG
```

```
SWLOKVKPELLVTFKGFQVSGAAVLCSSTVADYLLQFARHLIYSTMPPAQOAL
RASLAVIRSEGDARREKLAALITFRAGVQDLPTLADSCSAIOPLIVGDNRAQLQ
AEKLROOQGWTAIRPPTVPAGTARLRLTLTAHEMQDIDRLLEVLHGNG"
3030..3045
/standard_name="BIOD RBS"
3043..3753
/gene="BIOD"
3043..3753
/gene="BIOD"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYFTVGTDTVEGKTVASCALLQAAKAAGYRTAGYKPVASGS
EKTPGILNSDALALORNSSLOLDYATVNPYFAEPTSPHIIISAQEGRPILSVMSAG
LRALQQADWLVVEGAGGWETPLSDYATVFPADWVTEQOLPVLVVGKLGICINHAMLTA
QVIOHAGLTLAGWVANDVTPPKRHAHYMTLTRMPAPLLGELPWLAEHPENAATGK
YINLAFVDASTLGTFTSL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgctcgaagtcacagaatta 25
|||||
Db 137 GACATTGTCGCAAGTCACAGAATTA 161

RESULT 8
A93674 5872 bp DNA PAT 22-JAN-2000
LOCUS
DEFINITION Sequence 1 from Patent EP0798384.
ACCESSION A93674
VERSION A93674.1 GI:6741862
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia
1 (bases 1 to 5872)
Birch,O. and Brass,J.
Biotechnological method of producing biotin
Patent: EP 0798384-A 1 01-OCT-1997;
LONZA AG (CH)
Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A-15/9"
1..96
/function="PROMOTER PTAC"
/evidence=experimental
23..28
/standard_name="PROMOTER PTAC"
/evidence=experimental
45..49
/standard_name="BIOD RBS NO.9"
105..119
/evidence=experimental
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"
/codon_start=1

-35_signal
-10_signal

RBS
gene
CDS

FEATURES
source
Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAA02329.1"
/db_xref="GI:2294850"
/translation="MSWQEKINAAIDARRADALRRYPVAQAGRWLVADDRQYLNLF
SSNDYLGUSHHPQIRAWQQAQGFGLGSGSGHVSQYVYVHQAEELEAENLQYSEA
LFTSGFAANQAVTAAAKEDRIAADRSHASLLEAAKLSPSQLRRFAHNDVHLAR
LLASPCQGMVVTVEGFSMDGDSAPLAEITQQVTHQHWLMDVDDAHGTGVIGQGRG
```



YINLAFVDASTLGTSTRL"  
 BASE COUNT 1318 a 1552 c 1695 g 1307 t  
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;  
 Best Local Similarity 100.0%; Pred. No. 0.043;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaatta 25  
 |||||  
 Db 137 GACATTGTCGCAAGTCACAGAATTA 161

## RESULT 10

AR101809  
 LOCUS AR101809 5872 bp DNA PAT 14-FEB-2001  
 DEFINITION Sequence 1 from patent US 6083712.  
 ACCESSION AR101809  
 VERSION AR101809.1 GI:12812607  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 5872)  
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.  
 TITLE Biotechnological method of producing biotin  
 JOURNAL Patent: US 6083712-A 1 04-JUL-2000;  
 FEATURES Location/Qualifiers  
 source 1..5872  
 /organism="unknown"

BASE COUNT 1318 a 1552 c 1695 g 1307 t  
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;  
 Best Local Similarity 100.0%; Pred. No. 0.043;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaatta 25  
 |||||  
 Db 137 GACATTGTCGCAAGTCACAGAATTA 161

## RESULT 11

AR101810  
 LOCUS AR101810 5872 bp DNA PAT 14-FEB-2001  
 DEFINITION Sequence 6 from patent US 6083712.  
 ACCESSION AR101810  
 VERSION AR101810.1 GI:12812608  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 5872)  
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.  
 TITLE Biotechnological method of producing biotin  
 JOURNAL Patent: US 6083712-A 6 04-JUL-2000;  
 FEATURES Location/Qualifiers  
 source 1..5872  
 /organism="unknown"

BASE COUNT 1318 a 1552 c 1695 g 1307 t  
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;  
 Best Local Similarity 100.0%; Pred. No. 0.043;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaatta 25  
 |||||  
 Db 137 GACATTGTCGCAAGTCACAGAATTA 161

RESULT 12  
 AE000180  
 LOCUS DEFINITION

Escherichia coli K12 MG1655 section 70 of 400 of the complete genome.

ACCESSION AE000180 U00096  
 VERSION AE000180.1 GI:1786988  
 KEYWORDS

SOURCE ORGANISM

Escherichia coli K12.  
 Escherichia coli K12  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Escherichia

REFERENCE 1 (bases 1 to 11022)

AUTHORS

Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,  
 Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,  
 Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,  
 Mau, B. and Shao, Y.

The complete genome sequence of Escherichia coli K-12

Science 277 (5331), 1453-1474 (1997)

97426617

9278503

REFERENCE 2 (bases 1 to 11022)

AUTHORS

TITLE

Direct Submission  
 Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,  
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:  
 608-263-7459

REFERENCE 3 (bases 1 to 11022)

AUTHORS

TITLE

Direct Submission  
 Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,  
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:  
 608-263-7459

REFERENCE 4 (bases 1 to 11022)

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  
 This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG0301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). \*\*\* The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES  
 source

Location/Qualifiers  
 1..11022  
 /organism="Escherichia coli K12"  
 /strain="K12"  
 /sub\_strain="MG1655"

```

/db_xref="taxon:83333"
<1..130
/feature="REP (repetitive extragenic palindromic) element;
contains 4 REP sequences"
complement(147..1430)
/gene="ybhc"
/note="b0772"
complement(147..1430)
/gene="ybhc"
/function="orf; Not classified"
/note="f427; 98 pct identical to fragment YBHC_ECOLI
SW:P46130 (300 aa) but contains 127 additional C-terminal
residues"
/codon_start=1
/transl_table=11
/product="putative pectinesterase"
/protein_id="AAC73859.1"
/db_xref="GI:1786980"
/translation="MNTFSRSLALALAFVTLTACSTPPDQRPDSQDTAGTSSRPI
LSAKEAQNFAQHYFSLTFGAANPNPSPITLPAQDFVYVGPAGTQGVHTTIOAAVD
AAIIRKTKROYIAVMEGYOGTVYVPAAPGGITLYGTEKPIDVKIGLSLDCGMSPA
DWRHDVNPGRKGYMPKPAWYDSCQSKSDSIVLCSAVFWSQNNGLQLOLNTIENT
LGSVDAGNHPAVALRTDQDOVINNVNIIILGRONTPEVTNSGVQNRLETNROPRTLV
NSYIEGDVIVSGRAVFONTFERVNVSETOQEAYPEAPATLSNIYGFGLAVNSRFN
AFGQVAGLGRSLVDANTNGQVIRDSALNEGFNTAKPWADAVISNRPAGTGSVD
DNDEIQRLNDTINRWENNRNCGVSKVVAEAKK"
complement(1507..1535)
/note="factor Sigma70; predicted +1 start at 806574"
complement(1582..2058)
/gene="ybhb"
/note="b0773"
complement(1582..2058)
/gene="ybhb"
/function="orf; Unknown"
/note="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC73860.1"
/db_xref="GI:1786990"
/translation="MKLISNDLRDGLKPHRVFNGMGYDGNISPHLAWDDVPAGTK
SYVTCYDPAITGSGWVHWVNLPAITRVLPQGFSGVLVAMPDGVLOTRTDFGKTG
YDGAAPKGTTHYIFTHALDIERIDVDGASGAMGVNFVHSLASAITAFMS"
complement(2108..2124)
/note="central position to predicted promoter:85.5"
bound_moiety="RnaS predicted site"
complement(2117..3406)
/gene="bioA"
/note="b0774"
complement(2117..3406)
/gene="bioA"
/EC_number="2.6.1.62"
/function="enzyme; Biosynthesis of cofactors, carriers;
Biotin"
/note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
/codon_start=1
/transl_table=11
/product="7,8-diaminopelargonic acid synthetase"
/protein_id="AAC73861.1"
/db_xref="GI:1786991"
/translation="MTTDDLAFQDRHIWHPYTSMTSPLPVYVVSAGCELIISDGRR
LVDGMSWAAIHGYNHPQLNAAKQIDAMSHVMFGGITHAPATELCRLKLVAMTPOP
LECVFLADSGVAVEAVMAKLOYQAKGEAROFRTFRNGYHGTDFGMSKVDPDNS
MHSVLKGYLPENLFAPOQSDMGWDEROMVGFARLMAHSHETAATVIEPIVOGAG
GMRVYHPWLKRRKICDREGILLIADETATGFGRTGKLFACHEAIEADIILCLGKAL
TGTWTLSATLTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESQWQQ
QVADIEVQUREAPARDAEMADVRLVGAIVGVETTHPVNMAALQKFFVEQGVWIRP
FGKLIYMPPIITLPQQLRLTAANRAVQDETFECQ"
complement(2193..2221)
/gene="bioA"
/note="factor Sigma70; predicted +1 start at 807260"
3411..3450
/note="central position to bioB promoter: -20"

```

---

```

protein_bind
/db_moiety="BioB documented site"
3411..3450
/note="central position to predicted promoter: -20; genetic
evidence for the site"
bound_moiety="BioB predicted site"
complement(3411..3450)
/note="central position to predicted promoter:50"
bound_moiety="BioB predicted site"
complement(3411..3450)
/note="central position to bioA promoter:50"
bound_moiety="BioB documented site"
3413..3441
/note="factor Sigma70; promoter bioB; documented +1 at
808525"
complement(3447..3473)
/note="factor Sigma70; promoter bioA; documented +1 at
808515"
3493..4533
/gene="bioB"
/note="b0775"
3493..4533
/gene="bioB"
/EC_number="2.8.1.-"
/function="enzyme; Biosynthesis of cofactors, carriers;
Biotin"
/note="Q346; 99 pct identical to BIOB_ECOLI SW: P12996"
/codon_start=1
/product="biotin synthesis, sulfur insertion?"
/protein_id="AAC73862.1"
/db_xref="GI:1786992"
/translation="MAHRPWTLSQVTELFEPKPLDLLFEAQOVHROHFDPRQVQVST
LLSIKTGACPEDCKYQPSRYKTGLEARLMEVEQVLESARAKAAGSTRFCMGAAM
KNPDRMPYLEQVQVQKAMGLEACMTLGTLSAQOARLANAGLDYVNNHLDTSPEF
YGNIIITTYOERLDTLEKVRDAGTKVCGSGGIVGLGETVKDRAGLLQLANLPTPES
VPINLVKVGTPLADNDVDADFETIRIARVIMPTSYVRLSAGREQMNQEQAMC
FMAGANSIFYGCKLLITTPNEEDKDLQFRKLGLNPQQTAVLAGDNEQQQRLQALMT
PDTDEYYNAAAL"
4530..5684
/gene="bioF"
/note="b0776"
4530..5684
/gene="bioF"
/EC_number="2.3.1.47"
/function="enzyme; Biosynthesis of cofactors, carriers;
Biotin"
/note="Q384; 100 pct identical to BIOF_ECOLI SW: P12998"
/codon_start=1
/transl_table=11
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAC73863.1"
/db_xref="GI:1786993"
/translation="MSWQEKINAALDARRADARRRYPVQAQAGRWLVADDRQYLNLF
SSNDYLGLSHHPQITIRAWQOGEQIGICSGSHVSGYSVYHQALEBELAEWLGYRA
LLFTSGFAANOAVIAAMAKEDRIAADRLSHASLEAASLSPSOLRFAHNDVTHLAR
LLASPCQCOMVTEGVFSDGDSAPLAETQQVTOQHNGLMVDADHGTGTVGEQGRG
SCWLQVKPELLVTFKGFVGSAAVLCSSVTDVADYLLQFARHLIYSTMPPQAQAL"

```

---

```

Query Match 100.0%; Score 25; DB 1; Length 11022;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaatta 25
|||||
Db 3513 GACATTGTGCGCAAGTCACAGAATTA 3537

RESULT 13
AE005258 13501 bp DNA BCT 21-MAR-2001
LOCUS Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
DEFINITION of 155.
ACCESSION AE005258 AE005174

```

```

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postfai,G., Hackett,J., Link,S., Boutin,A., Shao,Y., Miller,I.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
TITLE
JOURNAL
MEDLINE
21074935
PUBMED
11206551
REFERENCE
AUTHORS
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postfai,G., Hackett,J., Link,S., Boutin,A., Shao,Y., Miller,I.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
TITLE
JOURNAL
MEDLINE
21074935
PUBMED
11206551
FEATURES
source
1. .13501
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohemorrhagic"
<1. .7576
/note="O-island #36; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655; Cryptic prophage
CP-933K; includes one copy of the 13 bp direct repeat that
flanks the prophage"
66. .665
/gene="lomk"
/note="lomk"
66. .665
/gene="lomk"
/function="putative membrane; Other or unknown (Phage or
Prophage Related)"
/note="Residues 1 to 199 of 199 are 72.86 pct identical to
residues 1 to 199 of 199 from GenPept 118 :
gi|7532789|gb|AAF63231.1|AF151091_2 (AF151091) Lom
[prophage P-Eida]"
/codon_start=1
/transl_table=11
/product="putative outer membrane protein of prophage
CP-933K"
/protein_id="AAG55137.1"
/db_xref="GI:12513752"
/translation="MRKVCALISLAICLVSGVPAWASEHQSTLSAGYLHASTDAPG
SDDLNGINVKYREFYTLGLTSFSYANAEDEQKTHYSDRWHEDYVNRWFVYVWAG
PSYRVNWFVYVWAGVAYSRVSTFSGDYFRVDRKRKTHVLTGSDDARYSNTSLAW
GAGVQPNPTESVAVDVAYEXSGDWRDTGFTVGVGYKF"
719. .2041
/gene="Z0982"
719. .2041
/gene="Z0982"
/function="putative structure; Structural component (Phage
or Prophage Related)"
/note="Residues 164 to 440 of 440 are 68.79 pct identical
to residues 381 to 645 of 645 from GenPept 118 :
gi|4585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
tail fiber protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="putative tail component of prophage CP-933K"

/protein_id="AAG55138.1"
/db_xref="GI:12513753"
/translation="MGMAAVQISGVKLDGAGKPIQNTIQLKARNSTXVYVNTVASE
NPDEAGRYSMDEYGYQYSVTLVVEGPPSHAGTISVYEDQSGPTLNDLFGAMTDDAR
PEARLRFQEEAAKAEAEAKNAGEAETASRNAGISAKAEASAAANADTAEDASE
SARQAAEASAAKSEASAAQKASQESLQATDAELSKKTAESAAGNAARDA
TTSTKARSAESAQSAEQSRIAEDAENVRIPTVVGPPGPKGPPGAGPQKGGKQGA
RGDTGATGATGEGPGGDTGPGAGPOGPKDGRGERTGLTGNAGPOGPGKDTGAAGPA
GPOGKGTGAAGVATGPOGPKGDFGETQIFRGLPMRLIETNSYSGFPFGTGDGALI
TGTLFDLPDKATQVQGFQHLQVRFQGGPQWDVKGDLDEVGSDTGRTE"
2004. .2312
/gene="Z0984"
2004. .2312
/gene="Z0984"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 102 of 102 are 98.03 pct identical to
residues 28 to 129 of 129 from GenPept 118 :
gi|4585437|gb|AAD25465.1|AF125520_60 (AF125520)
hypothetical protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55139.1"
/db_xref="GI:12513754"
/translation="MKWAVIQAELENDMNLRLKMQSLGCGKHDDCENGSRSLTAQLR
LGPADILDESDENGIIIPQDRVITQVVILDADKKQIQCVVRPIQLRADGTWENIGMK"
2489. .3469
/gene="Z0985"
2489. .3469
/gene="Z0985"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 13 to 326 of 326 are 55.73 pct identical
to residues 13 to 331 of 336 from GenPept 118 :
gi|6960367|gb|AAF33527.1| (AF170176) hypothetical protein
predicted by Glimmer [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/translation="MLSPITRTHNSVNIQSSPCQTVSFAGKEYELKVIDEKPILF
QWEPNPERYKKDEVPIVNTKQHPYLDNVTNAARIESDRMIGFVDCGDFSVNQKTAFS
KLEDFENVMILYREDYDFSMYDRKLSDIYDHIICEORLTEDKRDYELNLLKEULR
EISKAQSLISMAYAKKRNHAFDFRNALLKAGEIIFCTYNTKNHGCISFEGECIYLD
MDMLTLKGLTIYAPDGLISMHVDRRNDVNIENSALIVNKNHPALLEGSLFMSHKYD
AHPIYDGLGKGVKKYFNFTPLHNYNHFCDIEFNHPNLIIMNTSOYTCSSW"
3530. .4522
/gene="Z0986"
3530. .4522
/gene="Z0986"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/translation="MKPIPSLQSNFNSAPAGYSAPIAPNRAENAYADYVLDIGKRIPL
SAADLSNVYESVIRAVHDSRSLIDQHTVDMIGNTVLDALRSQTFDSDVSYGTHNEK
VHICIKYNEVELNEESSVKIDDIQSLTCNELYDYVQOEPIFPICEAGENDNEEYP
VSFVAPDSDSYEMPSWQEGLIHELIIHVHTGSSDPSGSDNTELGPTETELARRVAQELG
WSVPDFKGYAEPEAEHLRLNLRNALRQAMRHEENRAPERLGTISDRYEASPDFT
EYSVSNVTGYGFIQOHDPFGALINDINQDANQIQLYHGAPYIFTFGDVKHNNQ"
5350. .6231
/gene="Z0989"
5350. .6231
/gene="Z0989"
/function="orf; Other or unknown (Phage or Prophage
Related)"

```

```

/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55142.1"
/db_xref="GI:12513757"
/translation="MLSPYSNVLGCSWNSLTRLNLTSPDNRLVSSVRDAVHSDNGAQM
KVGNTYRVATDNKFCVTRSHSGCFTNLLHRLGWPKEISRKIEVYMLNAPVSAAM
ERGIVHSNRPLDPVDPVAPPPLPSVYDNRVLSVPCNVLGKGNVAVYEDADAQVLMK
FTTSQNEVTSVRCNOYXGASAEKIYVGNNGDIIGIRMDKINGESLINSLSLAQ
AEHALYDMEDRLQKGLIFVDVTTETNVLVDRANKNEFNPDISSNVNSWSSESQIMQ
SYHGKQDLISVLSKI"
/gene="7160"
/gene="z0990"
/gene="z0990"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55143.1"
/db_xref="GI:12513758"
/translation="MRPTSLNVLHQSSRSSMSDDTDIESLVKASSVQVTKNNPQLRF
QGDHNIYQIEAALDKIGSTETGRLVLLNAIESIRLKSSTVYVHLSNRLGVMAHRD
IDAENHRGSDGPHCNNAVYPCGEGISVYDFHATVTFHELLHVFHNLGRLKVES
SRAESQYSPILLLEARTVGLGAFSEVLSENFHEIGMPRTSYPPXDSALIHDDNT
VSLGFOOVRHLPLL"
/complement(7668. .8144)
/gene="ybhb"
/gene="z0992"
/complement(7668. .8144)
/gene="ybhb"
/function="orf; Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655; B0773"
/codon_start=1
/transl_table=11
/product="orf, Hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGLPDRHVFNGMGVYDGNISPHLAWDDVPAGTK
SFVYTCVDPDAPGSGWHVWVYVNLADTRVLDPQFGSLVAMPDGLVQTRIDFGKGT
YDGAAPKGGTHRIFTVHALDVERIDVDGASGAMVGFNVHPSLASASITAMFS"
/complement(8203. .9492)
/gene="bioA"
/note="z0993"
/complement(8203. .9492)

Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 qacattgtcgcaagtcacagaatta 25
|||||
Db 9599 GACATTGTCGCAAGTCACAGAAATTA 9623

RESULT 14
AP002553
LOCUS
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION AP002553 BA000007
VERSION AP002553.1 GI:13360211
KEYWORDS
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (sites)

```

```

AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
COMMENT
FEATURES
source
gene
CDS

```

Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S., Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T., Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T., Sakakawa, C. and Shinagawa, H.

Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic *Escherichia coli* O157:H7 derived from the Sakai outbreak

Genes Genet. Syst. 74 (5), 227-239 (1999)

20198780

2 (sites)

Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M., Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and Hayashi, T.

Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic *Escherichia coli* O157:H7 strain and an *Escherichia coli* K-12 strain MG1655

Syst. Appl. Microbiol. 23 (3), 315-324 (2000)

20557356

3 (sites)

Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sakakawa, C. and Shinagawa, H.

Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic *Escherichia coli* O157:H7 strain derived from the Sakai outbreak

Gene 258 (1-2), 127-139 (2000)

20564182

4 (sites)

Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sakakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and Shinagawa, H.

Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12

DNA Res. 8 (1), 11-22 (2001)

21156231

5 (bases 1 to 297816)

Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and Hayashi, T.

Direct Submission

Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: ken@gen-info.osaka-u.ac.jp, URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365, Fax: 81-6-6879-2047)

genome project.

Location/Qualifiers

1..297816

/organism="Escherichia coli O157:H7"

/strain="O157:H7"

/sub\_strain="RIMD 0509952"

/db\_xref="taxon:83334"

79..1245

/gene="ECs0753"

79..1245

/gene="ECs0753"

/note="similar to SUCC\_ECOLI gi1786948 percent identity 100 in 388 aa (Conserved in E.coli K-12)"

/codon\_start=1

/transl\_table=11

/evidence=not\_experimental

/product="succinyl-CoA synthetase beta subunit"

/protein\_id="BAB34176.1"

/db\_xref="GI:13360212"

/translation="MNLHEYQAKOLFAPYGLPAPVGYACTTPREAEASKIGAPWVVKOVHAGRGKAGVKVNSKEDIRAFENWLGKRLVYQTDANGQPNQILVEAATDIKALYLGAVDRSSRRVFMASSTGGVEIEKVAETPLIHKVALDPLTGPYOGRELAFKLGGLGVQVQTKIFMGLATLFLERDLALIEINPLVTITKQGDLCIDGKIGA DGNLFPQDLREMRQSDPREAQAWELNVALDGNIGCVANGAGLAMGTMDIV KLHGGEPAFLDVGGAETKRAFKIILSDDKVRVLVNIFFGTVRCDLIADGIIG AVAEGVNVPPVVRLEGNNAELGAKKALDADSLNIIAAKGLTDAQQVVAAVEK"



```

1245. .2114
/ gene="ECs0754"
CDS
/ gene="ECs0754"
/ note="similar to SUCD_ECOLI gi11786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/ codon_start=1
/ transl_table=11
/ evidence=not_experimental
/ product="succinyl-CoA synthetase alpha subunit"
/ protein_id="BAB34177.1"
/ db_xref="GI:13360213"
/ translation="MSLIDNKNKVICQGTGSGQTFHSEQAATAYGATKXVGGVTPGKG
GTHLGLPNTVREAAVATGASVIYVPAPEKDSILEAIDAGIKLIITTEGPT
LDMLTVKLDAGVREKNCPCVITPGECKIGIQGHFKPGKGVIVSRSTLTVE
AVKOTDYGQSCVIGIGDPGPSNFIDILEMFKDPQTEAIVMIGETGGSAEEA
AAYTKEHVTVPVGYIAGVTAPKGRMHAGAIAGGKGTADKFAALEAAGVKTTRS
LADIGEALKYVLK"
/ gene="ECs0755"
CDS
/ gene="ECs0755"
/ note="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator) -
Escherichia coli gi1417043|sp|P32064|GCVA_ECOLI percent
identity 31 in 300 aa"
/ codon_start=1
/ transl_table=11
/ evidence=not_experimental
/ product="putative transcription regulator"
/ protein_id="BAB34178.1"
/ db_xref="GI:13360214"
/ translation="MRGKIPKTELLVTFEVVARHESYTRAAELALQSAVFRQVSAL
REFLTPFSKRRIFLNDGKYVIGVIGVETLNKLERDNTIMTQPTVQVIELAVN
PTESPHLIPNLHETKLHPDI IVNIHSLANGDFLNREYDAVIMRENFCAPAEVEY
LFEBELLPCVSGSLLMSDKLSVAELLTELPLHQSPTITGWEWELSCVSSPLVN
NGPRDLLSLIAVRSNLGVALLPRAIQLHDLSDGMVPCDVPITRGNFIMTWQE
EKSDSPHLQOQFRENLLAKSVVPOEM"
/ complement(3720. .4322)
/ gene="ECs0756"
CDS
/ complement(3720. .4322)
/ gene="ECs0756"
/ note="probable cob(I)alamin adenosyltransferase, similar
to cob(I)alamin adenosyltransferases (corrinoid
adenosyltransferases) e.g. [Escherichia coli]
gi115148|sp|P13040|BTUR_ECOLI percent identity 67 in 200
aa"
/ codon_start=1
/ transl_table=11
/ evidence=not_experimental
/ product="putative cob(I)alamin adenosyltransferase"
/ protein_id="BAB34179.1"
/ db_xref="GI:13360215"
/ translation="NEARTSTERHVORQKLKEQVDTRVAAAIEKKGILLIVFTGNGKG
KSTAFGTVTRAVGKGVQVQYIKGQMDNGEYNLLOPLGVFERHIMGTGFTWETNR
QADIDAAKVESKRLAKDRKLDLWLDELTYMLAYHLDTEEVIASLQNRPAQOSV
IVTGRCHSQILKMDTVSEIRPVKHAFDNGIQAPGIDW"
/ complement(4332. .5984)
/ gene="ECs0757"
CDS
/ complement(4332. .5984)
/ gene="ECs0757"
/ note="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarase) - Escherichia coli
gi120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545
aa"
/ codon_start=1
/ transl_table=11
/ evidence=not_experimental
/ product="putative fumarate hydratase"
/ protein_id="BAB34180.1"
/ db_xref="GI:13360216"

/translation="MSKPFITWQEPFLONKDCGYTTLISDQHITVTELDGEEVIAKPAE
ALTLARQAFYEASFLSAHQVAVSILNDQOASSNDKYVALOLLNNAEYSAKGVLP
NCODGTATIVASKGQOIWTGNDAAEALSKGIYSTFOENLRFSONAPLDMYTVETQ
TNLPAQIDISAVAGDEYHFLCVNKGGSANRAALYQETKSLIQPEKLTAFLEKMSL
GTAACPYPYIAFVVGSLADQTLKIAKLASTKYNDLPTSGNEOQAFKQFLEKYL
EASQOFGICAGFGKYFAHDIRVIRLPRHGGSCPIAMALSCSADRNTKAKINKHGILW
EKLEHNPQYIPASLREYHQAQVOLDLNRPLRDVMDLARLPVGTIVSLSGPIVVAR
DIAHAKIKARLDSGPMPEYLKHHIVYVAGPAKTPENMACGSLGPTTGGMDGYIDTF
QAAGSLWLSKGNRSQOVDACHKHGGFNLSIGSAAALAAQEVVKSRLCLEYPELG
MEAVMMEVENLPAILVDDKGNFFSQEQHRCASCAPAH"
/ complement(6092. .7372)
/ gene="ECs0758"
CDS
/ complement(6092. .7372)
/ gene="ECs0758"
/ note="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
gi1121467|sp|P24943|GLTT_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transpor"
/ codon_start=1
/ transl_table=11
/ evidence=not_experimental
/ product="putative transport protein"
/ protein_id="BAB34181.1"
/ db_xref="GI:13360217"
/ translation="MKKISLITMILALVLGMLIIVLNNNTASPETAKLYAQEISIFT
TIFRLIKMIIAPLVSTLVGTAKMGDAKALGRIFSKTFLFICASLSIALGLITY
NFFMGTGINFVAHGAETGVAAEPTLKVELSHAFPTSIDVMAHNEILOIVVFSI
FJGCSITAIGKSGSAIVHALDSLAHAKLTGTVMFLFAPLTFASALIAERGLAVM
VSAGIFMGFEYFTMLLVLGLIAIVYVGPICIRRLTRALSEPALLIASITSEARFP
GTLEKLEQFGVSPKIASFVLPYIGSYFNLSMAYCSFATVITACQNIHLSIQEITM
LLTLMLTSGMAGVPRASVVIATLQNFNPEAGLILMLGVDPFLDMGRSATNVMSN
AMGAAMVSWEGEHGEGCGRKALKPNESNALP"
/ complement(7533. .7850)
/ gene="ECs0759"
CDS
/ complement(7533. .7850)
/ gene="ECs0759"
/ note="unknown"
/ codon_start=1
/ transl_table=11
/ evidence=not_experimental
/ product="hypothetical protein"
/ protein_id="BAB34182.1"

Query Match 100.0%; Score 25; DB 2; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.042;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaatta 25
|||||
Db 92331 GACATTGTCGCAAGTCACAGAATTA 92355

RESULT 15
AF250776 5526 bp DNA BCT 31-JAN-2001
LOCUS uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
DEFINITION modC-bioA intergenic region, DAPA-aminotransferase bioA (bioA),
biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
biotin biosynthesis protein BioC (bioC) genes, complete cds; and
dethiobiotin synthetase BioD (bioD) gene, partial cds.
ACCESSION AF250776
VERSION AF250776.1 GI:12620124
KEYWORDS uncultured bacterium pCosHE2.
SOURCE uncultured bacterium pCosHE2
ORGANISM Bacteria; environmental samples.
REFERENCE 1 (bases 1 to 5526)
AUTHORS Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streit, W.R.
TITLE Direct cloning from enrichment cultures, a reliable strategy for
isolating complete operons and genes from microbial consortia
JOURNAL Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
MEDLINE 20575196

```

```

PUBMED 11133432
REFERENCE 2 (bases 1 to 5526)
AUTHORS Entcheva,P., Liebl,W. and Streit,W.R.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES
source
1. .5526
/organism="uncultured bacterium pCosHE2"
/db_xref="taxon:143797"
/clone="pCosHE2"
/note="unknown organism, cosmid clone derived from
environmental consortium"
/complement(52. .528)
/note="ORF1"
/codon_start=1
/transl_table=11
/product="hypothetical 17.1 kDa protein in modC-bioA
intergenic region"
/protein_id="AAG60577.1"
/db_xref="GI:12620125"
/translation="MKLISNDLRDGDKLPHRVFMGMYDGDNTSPHLAWDDVPAGTK
SEVVCYDPAFTGSGWVWVNLPAOTRVLPQFGSLVAMPDGVLTQPTDFGKAG
YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGMVGFVHFSLASASITAMFS"
/complement(587. .1876)
/gene="bioA"
/complement(587. .1876)
/gene="bioA"
/note="7,8-diaminopelargonic acid
synthetase-aminotransferase"
/codon_start=1
/transl_table=11
/product="DAPA-aminotransferase BioA"
/protein_id="AAG60578.1"
/db_xref="GI:12620126"
/translation="MTTDLAFDQDRHIWHPYTSPLPVYVVSAGCEILISDGRR
LVDCSSWAAIHGYNHNPOLNAAMKSOIDAMSHVYFGIITHAPALELCRLKIVAMPQP
LKFVFLADPSGVAVEMAKALQYQAKGEARQELFRNGYHGDFTGAMSCVDPDMS
MHLWKGVLPLNLPAPQPSRMDGEWDMVGFARLMAHRHEIAAVIIEPIVQVQAG
GMWYHPWLRKIRKIDREGILLIADIETAFGTGRTKLFACAEHATPDIICIGKAG
TGGTWLSATLTREVAETISNGEAGCFMGFTMGPNPLACAAANSLAILESGDMQO
QVADIEVLRQLAPARDAEMADVRLGAVVETHPVNMALQKFFVEVQGVWIRP
FGKLIYLPYPIYLLPQOLRLTAANRAVDQETFFCQ"
1963. .3003
/gene="bioB"
1963. .3003
/gene="bioB"
/codon_start=1
/transl_table=11
/product="biotin synthase BioB"
/protein_id="AAG60579.1"
/db_xref="GI:12620127"
/translation="MAHRPWTLSQVTELFKPLDLLLFFAQOVHROHEDPROQVST
LLSITGACPEDCKYCPQSSRYKTGLEAERLMEVQVLESARKAKAAGSTRFCMGAAW
KNPNERDMPYLEOVGVKDLGLEACMTGLTSEQAORLANAGLDYNNHNDISPEF
YGNLITFRYQRLDLEKVRDAGIKVCSGGIVGLGETVKDRAGLLQLLANLPTPES
VPINMLYKVTGLAONDVDADFRTIIVARIMPTSYVRLSAGREOMNEQTQMC
FMAGANSIFYGCKLLTPNPEEDKDLQFKRLGLNLPQQTAVLAGDNEQQORLEQALMT
PDTDEVYNAAL"
3000. .4154
/gene="bioF"
3000. .4154
/gene="bioF"
/note="8-amino-7-oxononanoate synthase"
/codon_start=1
/transl_table=11
/product="KAPA synthetase BioF"
/protein_id="AAG60580.1"
/db_xref="GI:12620128"
/translation="MSWQEKINAALDARRAADALRRRYPVAQAGRWLWADROVLNF
SSNDYLGSHHPQIIRAWQOGEAFGIGSGSGSHVSVYVHQAEELEAEVLYVNF
LFIISGFAANOAVIAAMAKEDRIADRLSLASLEAASLSQSLRRFAHNDVTHLAR
LLASPCGGQQLVTEGVFMSDGSAPLAEIQVTOHNGWLMVDDAHGTGVTGEGRG
SCWLQVKVPELLVVTFCGKGVSGAAVLCSTVADYLLQFARHLIYSTSMPPQAQAL
RASLAVIRROGDARREKLVSLIAHFRAGVQDLPFTLADSCSAIQLIVGDNRAQLQ
AEKLRQOCGWVTAIRPPTVPAGTARLRLTLTAHEMODIDRLLEVLHGNC"
4141. .4896
/gene="bioC"
4141. .4896
/gene="bioC"
/note="reaction step prior to synthesis of pimeloyl-CoA"
/codon_start=1
/transl_table=11
/product="biotin biosynthesis protein BioC"
/protein_id="AAG60581.1"
/db_xref="GI:12620129"
/translation="MATVKNQAJAAAFRAAALYEQHADLQROASDVLLAMLPERKYT
HVIDAGCGPMWTRHMRHAQVLTALDSPMLVOARQDAADHYLAGDTESLPLATA
TFDLANSNAVWCMGNLSALRELYRVRSKGVAFVTLVQGSPLPELHOAWAVDERP
HANRFLPPDEIEQSLANGVYQHHTIQTITLWFDLALSAMRSKLGIGATHLHEGRDPRIL
TRSQLORLQAWPQQQGRVPLTYHLFLGVARE"
4883. .5526
/gene="bioD"
4883. .5526
/gene="bioD"
/note="DTB synthetase"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase BioD"
/protein_id="AAG60582.1"
/db_xref="GI:12620130"
/translation="MLVSKRYFVTGDTVEGKTVASCALLOAKAAGYTAGYKPVAS
GSEKTEGLRNSDALQORNSLDYATVNPYTFAPETSPHIIISAQGRPTESVMS
SGLRALEQADQWLVFEGAGGWFPISDTFTFADWVTOEOLPVLVGVKLGICINHAML
TAQAIQHAGITLAGVANDVTPGKRHAIEYITLIRMIAPLLGEIPWLAENPE"
BASE COUNT 1274 a 1507 c 1567 g 1178 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 2; Length 5526;
Best Local Similarity 96.0%; Pred. No. 0.29;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacattgtcgaagtacagaatta 25
|||||
Db 1983 GACCTGTGCGAAGTACAGAATTA 2007

RESULT 16
AF248314
LOCUS
DEFINITION
AF248314 8227 bp DNA BCT 24-JAN-2001
Uncultured bacterium pCosAS1 urocanase-like protein (hutU) gene,
partial cds; histidine ammonia-lyase-like protein (hutH), DAPA
aminotransferase BioA (bioA), biotin synthase BioB (bioB), 7-KAPA
synthetase (bioF), biotin biosynthesis BioC-like protein (bioC),
and dethiobiotin synthase BioD (bioD) genes, complete cds; ABC
transporter-like protein (elsa) gene, partial cds; and unknown
gene.
ACCESSION AF248314
VERSION AF248314.1 GI:12407610
KEYWORDS
SOURCE uncultured bacterium pCosAS1.
ORGANISM uncultured bacterium pCosAS1.
Bacteria; environmental samples.
REFERENCE 1 (bases 1 to 8227)
AUTHORS Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
TITLE Direct Cloning from Enrichment Cultures, a Reliable Strategy for
Isolation of Complete Operons and Genes from Microbial Consortia
Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
11133432
JOURNAL
PUBMED 2 (bases 1 to 8227)
AUTHORS Entcheva,P., Liebl,W. and Streit,W.R.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2000) Mikrobiologie und Genetik, Universitaet
Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES
Location/Qualifiers

```

```

source
1. .8227
/organism="uncultured bacterium pCosas1"
/db_xref="taxon:143793"
/clone="pCosas1"
/note="unknown organism, cosmid clone derived from
environmental consortium"
<1. .502
/gene="hut0"
<1. .502
/gene="hut0"
/note="putative urocanate hydratase"
/codon_start=2
/transl_table=11
/product="urocanase-like protein"
/protein_id="AAG53593.1"
/db_xref="GI:12407618"
/transl_table="PRERISQGLPARICWVLEWRQKLGIAFNEMVRRGEVSAPIVI
GRDLDSGVSAPNRETRAMODGSDAVSDWPLNALLNTASGATWVSLHGGGVGMGF
SQRAGMVICDGTDDAARIARVHLNDPATGVMRHADAGYDIAIDCAKEQGLNPLMIA
ATQGQH"
gene
502. .2031
/gene="hutH"
CDS
502. .2031
/gene="hutH"
/note="putative"
/codon_start=1
/transl_table=11
/product="histidine ammonia-lyase-like protein"
/protein_id="AAG53586.1"
/db_xref="GI:12407611"
/transl_table="MNALTTPCTTLAQLRQVMOQPLQLTIDESAHEAINDSVACVE
AIVAGRTAYGNTGFLGLAQTRIATHNLENSLRVLSHAAGVCEPLDDDLVRLMVY
LKINSIRGFSIRUSVIOALIALYNAGVISVDPKAGSGDGLAPLHSLMTLLGE
GKARYGELPATALQKAGLAPVTLAAKEGALLNGTQASTAFALRGLFAEDLFAS
AVVCALITTEAVLGRPPEDARIEHVRGQGGIDAAALFRHLVLTSAIESHNCCK
VDDPYSLRCQPOVMGACLTOMQROVAELLVESNDPLVFAAENEMVFRGNFAEP
VAMADNLATAETIGALSERTIALMDKHSQLPFLVRNGVNSGFMIAQVTAAL
ASENGLCHPHTSDVKPCPSANOEDHVSMAAGRLWEMAGNTRGVLAWEVLAC
QKADLRDGLTSPLLEQARQSCGEQVAHYDDRRFPADIEAAISLNLKSLGLLIPAF
L"
complement(2130..2606)
/note="ORF1; similar to hypothetical 17.1 kDa protein in
modC-bioA intergenic region"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG53587.1"
/db_xref="GI:12407612"
/transl_table="MKLISQDLRDGDKPLRLHVFNGMGVEGDNISPHLAWDRVPSGTK
SFVVTCTDPDPTDSGWHWIVANPADTRVLPGSGSLVALPEGAVETRTDFGKAG
YCGAAPKGETHRYIFTVHALNVERIEVDEEASGAMGVNFHFTLGSASITAMYS"
complement(2651..3940)
/gene="bioA"
complement(2651..3940)
/gene="bioA"
/note="7, 8-diamino-pelargonic acid aminotransferase"
/codon_start=1
/transl_table=11
/product="DAPA aminotransferase BioA"
/protein_id="AAG53588.1"
/db_xref="GI:12407613"
/transl_table="MTTDDLPDRHIWHPYTSMTOPLVPVPVPEAHGCELTLASGEQ
LVDGMSWAAIHGYNHRPLNRAKQIDRMSHVMGGITHOPAVDLCRLVAMTPET
LECVFPADSGSAVEVAMKALQYWHAKESQRQFITFRNGYHGDTFGMSVCDPDNS
MHSUWGYLPENLFAPOPSRQDGAEDMDVGLAPLMAAHRHEIAAVILEPIVQAG
GMRMYHPWLKRRMCDREGILLIADIEATGFGTKLFACEHAGITDILCLGKAL
TGGTMTMSATLTTRQVAETISNGEACFMHGPFTFMGNPLACAVASESLASLEQWQD
QVAIESOLEELAVARDSEFVADRVLGAIGVETHTPNNMAALQRFVFGVWVRP
EGKLIYLPVPVIRPDQLSKLTRAVDAVQAPAHFTR"
gene
4029. .5069
/gene="bioB"
CDS
4029. .5069
/gene="bioB"

/codon_start=1
/transl_table=11
/product="biotin synthase BioB"
/protein_id="AAG53589.1"
/db_xref="GI:12407614"
/transl_table="MAHQARWTLISQVTFALFKPLLELLFRAQOQIHRQHFDPQOIQVST
LLSTKTCACPDCKYQPOSARYTKGLESERLMEVEOVLESARQAKNAGSTRFQCMGAAM
KNPHERMPYLEQMOVGOKAMGLOACMTLGLTDHQAORLASAGLDYYNHNLDPPEF
YGNITRTYQERLDILDKVRDAGIKVCSSGIVGLGETYTDRAGLLLQLLANLPTPEES
VPINMLYKVKPTLADNDVXAFDXINTXSAXVAMAPTALXVLRSLSTREGQNEQTAQC
FMAGANSIFYGCKLLTTPNPEEDKQVLFKRLGNPQOATVLTGDNEDQHQLEQQLFN
ADTDQFYNAATV"
gene
5066. .6223
/gene="bioF"
CDS
5066. .6223
/gene="bioF"
/note="7-keto-8-amino-pelargonic acid synthetase; similar
to BioF"
/codon_start=1
/transl_table=11
/product="7-KAPA synthetase"
/protein_id="AAG53590.1"
/db_xref="GI:12407615"
/transl_table="MTWQQRIDAALDARRAADALRVTPVENGAGRFLIREQRRYCNF
SGNDYLGLSQHPETIRAWQOQAGQYGVGGSGHVSQYTHAQALBEHLAGWLGYPRA
LLFTSGFAANQALIAITDKREDRIVADRLSHASLLAESPAQLRRFAHNDVAOLDSD
LLDKPHGQQLVYVTEGIFSMGDSAPLGVIAHAAOQGSMLLVDDAHGIGVGTGEGRG
SAASGVKPELLVYTFKGVGVSAAVLCSDAVADYLLQFARHLIYSTSPPAQAVAL
LAAPNVRSEDNARRQALAIIDQFRAGVRDLPQVITASQSAIQPLVIGENARALHL
AQRRLRGIWVTIRPTVPAGTARLRLTLTAHQPADITALVEALHVAGE"
gene
6207. .6962
/gene="bioC"
CDS
6207. .6962
/gene="bioC"
/note="Biotin biosynthesis BioC-like protein"
/codon_start=1
/transl_table=11
/product="biotin biosynthesis BioC-like protein"
/protein_id="AAG53591.1"
/db_xref="GI:12407616"
/transl_table="MSQVKNQAIATAAFRAAQSYSRHDALOROSADGLRALLAENTFA
SVLDAGCGPSNLAWRQTSVVTAMDLCAPMLEEARRNQAAQSYLQADIEAIPLADA
QYSLAWSHLAVQXCASLPRAELMRVTQLGGKVAFTLLNGSLPELQADWQAVDTP
HANFLSQQVNDALSGMYRCAYQVTVTLFDDALSAMHSLKIGIGATHLHAGRAKPL
TRGQLQLAQAWPQOAGKFLPSYOLFHCIIERD"
gene
6940. .7680
/gene="bioD"
CDS
6940. .7680
/gene="bioD"
/note="dethiobiotin synthase BioD"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthase BioD"
/protein_id="AAG53592.1"
/db_xref="GI:12407617"
/transl_table="MELSNVTERFYVTGTEVCKTVASTALLOAAALGRHPTAGYKP
VASGEMTAQLRNSDALQNRNSLALSYPAVNPYTFAPETSPHISADAEQRIEFA
TSLAGLRALETQADWLVLEGGWFTPLSQFADWMAERLVLVIVGVKLGCIH
AMLTAQAIROAGLHFAGVANGVVPGRKHAEMYMATLRRVLSAPLLGEIPLWLGQADT
SAVGQYLDLTALTPTSAPSSGOSAGHPAR"
complement(7619..8227)
/gene="eisa"
complement(7619..8227)
/gene="eisa"
/note="putative urea/short-chain amide ABC transporter"
/codon_start=1
/transl_table=11
/product="ABC transporter-like protein"
/protein_id="AAG53594.1"
/db_xref="GI:12407619"
/transl_table="DCXILRTRXGNLXVSCIAGLPIASGTMMWHEAPGVNQAGTEI
XVVPQDRRIFSQLTVEENLHAQNAKGIAGSLSDVPEFPELPYLPQLRASGLND
QOYOLANALVTRPRLLILDEPSRGTGLFTVKLAQLLRNRELGLSVLLABQHLS
LIRRVADRFCLLYRGRNVAQGRVSELDPLIAHMGPTASQ"
gene
1512 a 2642 c 2605 g 1453 t 15 others
BASE COUNT

```

## ORIGIN

Query Match 87.2%; Score 21.8; DB 2; Length 8227;  
 Best Local Similarity 92.0%; Pred. No. 1.9;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 gacattgtcgcaagtccacagaatta 25  
 |||||  
 Db 4049 GACATTGTGCGCAAGTCACTGCATTA 4073

## RESULT 17

AC084797 172258 bp DNA HTG 17-NOV-2000  
 AC084797 Homo sapiens chromosome 16 clone RP11-6203, WORKING DRAFT SEQUENCE,  
 31 unordered pieces.  
 AC084797.1 GI:11192125  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 172258)  
 DOE Joint Genome Institute.

## AUTHORS

Sequencing of Human Chromosome 16

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 172258)  
 DOE Joint Genome Institute.

## AUTHORS

Direct Submission

## JOURNAL

Submitted (17-NOV-2000) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

## COMMENT

-----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: <http://www.jgi.doe.gov>  
 -----

## Project Information

Center Project Name: 0  
 Center clone name: RPCI-11\_6203  
 -----

## Summary Statistics

Consensus quality: 126064 bases at least Q40  
 Consensus quality: 142328 bases at least Q30  
 Consensus quality: 149310 bases at least Q20  
 Estimated insert size: 158300; agarose-fp estimation  
 Quality coverage: 4.23 in Q20 bases; sum-of-contigs estimation  
 Quality coverage: 3.95 in Q20 bases; sum-of-contigs estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 31 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1116: contig of 1116 bp in length  
 \* 1117 1216: gap of unknown length  
 \* 1217 2356: contig of 1140 bp in length  
 \* 2357 2456: gap of unknown length  
 \* 2457 3865: contig of 1409 bp in length  
 \* 3866 3965: gap of unknown length  
 \* 3966 5349: contig of 1384 bp in length  
 \* 5350 5449: gap of unknown length  
 \* 5450 6474: contig of 1025 bp in length  
 \* 6475 7846: gap of unknown length  
 \* 7847 7946: contig of 1272 bp in length  
 \* 7947 9123: contig of 1176 bp in length  
 \* 9123 9222: gap of unknown length  
 \* 9223 10296: contig of 1074 bp in length

\* 10297 10396: gap of unknown length  
 \* 10397 11704: contig of 1308 bp in length  
 \* 11705 11804: gap of unknown length  
 \* 11805 13371: contig of 1567 bp in length  
 \* 13372 13471: gap of unknown length  
 \* 13472 15810: contig of 2339 bp in length  
 \* 15811 15910: gap of unknown length  
 \* 15911 18530: contig of 2620 bp in length  
 \* 18531 18630: gap of unknown length  
 \* 18631 21218: contig of 2588 bp in length  
 \* 21219 21318: gap of unknown length  
 \* 21319 24015: contig of 2697 bp in length  
 \* 24016 24115: gap of unknown length  
 \* 24116 27354: contig of 3239 bp in length  
 \* 27355 27455: gap of unknown length  
 \* 27455 29518: contig of 2064 bp in length  
 \* 29519 29618: gap of unknown length  
 \* 29619 33178: contig of 3560 bp in length  
 \* 33179 33278: gap of unknown length  
 \* 33279 40853: contig of 7575 bp in length  
 \* 40854 40953: gap of unknown length  
 \* 40954 45631: contig of 4678 bp in length  
 \* 45632 45731: gap of unknown length  
 \* 45732 48140: contig of 2409 bp in length  
 \* 48141 48240: gap of unknown length  
 \* 48241 54342: contig of 6002 bp in length  
 \* 54343 54342: gap of unknown length  
 \* 54343 62528: contig of 8186 bp in length  
 \* 62529 62528: gap of unknown length  
 \* 62529 69663: contig of 7035 bp in length  
 \* 69664 69663: gap of unknown length  
 \* 69664 79925: contig of 10162 bp in length  
 \* 79926 80025: gap of unknown length  
 \* 80026 91695: contig of 11670 bp in length  
 \* 91696 91795: gap of unknown length  
 \* 91796 98537: contig of 6742 bp in length  
 \* 98538 98537: gap of unknown length  
 \* 98538 107062: contig of 8425 bp in length  
 \* 107063 107162: gap of unknown length  
 \* 107163 119401: contig of 12239 bp in length  
 \* 119402 119501: gap of unknown length  
 \* 119502 132164: contig of 12663 bp in length  
 \* 132165 132264: gap of unknown length  
 \* 132265 150238: contig of 17974 bp in length  
 \* 150239 150338: gap of unknown length  
 \* 150339 172258: contig of 21920 bp in length.

## FEATURES

Location/Qualifiers  
 1..172258  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /clone="RP11-6203"  
 /clone\_lib="RPCI human BAC library 11"

BASE COUNT 47622 a 36504 c 38345 g 45759 t 4028 Others  
 ORIGIN

Query Match 85.6%; Score 21.4; DB 77; Length 172258;  
 Best Local Similarity 95.7%; Pred. No. 3;  
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 acattgtcgcaagtccacagaatt 24  
 |||||

Db 58121 ACATTGTGCGCAAGTCACTGCAGATT 58143

## RESULT 18

LOCUS AF250770 965 bp DNA BCT 31-JAN-2001  
 DEFINITION Uncultured bacterium pCOSHE1 DAPA-aminotransferase (bioA) and  
 biotin synthase (bioB) genes, partial cds.  
 ACCESSION AF250770  
 VERSION AF250770.1 GI:12620104

```

KEYWORDS      uncultured bacterium pCosHE1.
SOURCE
ORGANISM      Bacteria; environmental samples.
REFERENCE     1 (bases 1 to 965)
AUTHORS       Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
TITLE         Direct cloning from enrichment cultures, a reliable strategy for
              isolation of complete operons and genes from microbial consortia
JOURNAL       Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
MEDLINE       20575196
PUBMED        11133432
REFERENCE     2 (bases 1 to 965)
AUTHORS       Entcheva,P., Liebl,W. and Streit,W.R.
TITLE         Direct Submission
JOURNAL       Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
              Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES      Location/Qualifiers
              1..965
               /organism="uncultured bacterium pCosHE1"
               /db_xref="taxon:143796"
               /clone="pCosHE1"
               /note="unknown organism, cosmid clone derived from
               environmental consortium"
               complement(<1..438)
               /gene="bioA"
               complement(<1..438)
               /gene="bioA"
               /note="7,8-diaminopelargonic acid
               synthetase-aminotransferase; similar to BioA"
               /codon_start=1
               /transl_table=11
               /product="DAPA-aminotransferase"
               /protein_id="AA60563.1"
               /db_xref="GI:12620105"
               /translation="MTQDDLAFDQHQHWHPTYSMTRPXPYPVVASAXAHCQRRRLVD
               GMSWAAIHGYNHRLNAAKQIQGMSHVFDGITHQPAVDLCRRXVAMTPDALEC
               VFLADSGVAXEVXENGAASTAPLPADRPERYHXAFIPXLQ"
               525..>965
               /gene="bioB"
               525..>965
               /gene="bioB"
               /note="similar to BioB"
               /codon_start=1
               /transl_table=11
               /product="biotin synthase"
               /protein_id="AA60564.1"
               /db_xref="GI:12620106"
               /translation="MXHHARWTMSQVTELFNKPFLFEMFAQQVHRQHFDPRHVQXST
               LLSTKGACDECKYCPQSARYKTGLESLRMEQVLDLSARKAKNAGSTRFCMGAAW
               KKPTIVTCPLYEQMKYKEMGLEALYDPRHERRAXSACLLA"
BASE COUNT    230 a 256 c 285 g 180 t 14 others
ORIGIN
Query Match      80.8%; Score 20.2; DB 2; Length 965;
Best Local Similarity 88.08; Pred. No. 13;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gacattgcgaagtacagaatta 25
    ||| ||||| ||||| ||||| |||||
Db 545 GACGATGTCGCAAGTCACTGAATTA 569

RESULT 19
AC090618/c      64231 bp      DNA      HTG      03-MAR-2001
LOCUS           Homo sapiens chromosome 17 clone RP11-763E3 map 17, LOW-PASS
DEFINITION      SEQUENCE SAMPLING.
ACCESSION       AC090618
VERSION          AC090618.1 GI:13194354
KEYWORDS         HTG; HTGS_PHASE0.
SOURCE           human.
ORGANISM         Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 64231)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone RP11-763E3
Unpublished
2 (bases 1 to 64231)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barra,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepei,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Glnde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., LaRoque,K., Lamazares,R., Landers,T.,
Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Riser,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12048
Center clone name: 763_E_3
-----
* NOTE: This record contains 79 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 632: contig of 632 bp in length
* 633 732: gap of 100 bp
* 733 1437: contig of 705 bp in length
* 1438 1537: gap of 100 bp
* 1538 2262: contig of 725 bp in length
* 2263 2362: gap of 100 bp
* 2363 3095: contig of 733 bp in length
* 3096 3195: gap of 100 bp
* 3196 3890: contig of 695 bp in length
* 3891 3990: gap of 100 bp
* 3991 4700: contig of 710 bp in length
* 4701 4800: gap of 100 bp
* 4801 5523: contig of 723 bp in length
* 5524 5623: gap of 100 bp
* 5624 6344: contig of 721 bp in length
* 6345 6444: gap of 100 bp
* 6445 7159: contig of 715 bp in length
* 7160 7259: gap of 100 bp
* 7260 7985: contig of 726 bp in length

```

```

* 7986 8085: gap of 100 bp
* 8086 8794: contig of 709 bp in length
* 8795 8894: gap of 100 bp
* 8895 9604: contig of 710 bp in length
* 9605 9704: gap of 100 bp
* 9705 10415: contig of 711 bp in length
* 10416 10515: gap of 100 bp
* 10516 11270: contig of 755 bp in length
* 11271 11370: gap of 100 bp
* 11371 12111: contig of 741 bp in length
* 12112 12211: gap of 100 bp
* 12212 12918: contig of 707 bp in length
* 12919 13018: gap of 100 bp
* 13019 13724: contig of 706 bp in length
* 13725 13824: gap of 100 bp
* 13825 14535: contig of 711 bp in length
* 14536 14635: gap of 100 bp
* 14636 15356: contig of 721 bp in length
* 15357 15456: gap of 100 bp
* 15457 16170: contig of 714 bp in length
* 16171 16270: gap of 100 bp
* 16271 16968: contig of 698 bp in length
* 16969 17068: gap of 100 bp
* 17069 17790: contig of 722 bp in length
* 17791 17890: gap of 100 bp
* 17891 18599: contig of 709 bp in length
* 18600 18699: gap of 100 bp
* 18700 19388: contig of 689 bp in length
* 19389 19488: gap of 100 bp
* 19489 20208: contig of 720 bp in length
* 20209 20308: gap of 100 bp
* 20309 21048: contig of 740 bp in length
* 21049 21148: gap of 100 bp
* 21149 21847: contig of 699 bp in length
* 21848 21947: gap of 100 bp
* 21948 22662: contig of 715 bp in length
* 22663 22762: gap of 100 bp
* 22763 23479: contig of 717 bp in length
* 23480 23579: gap of 100 bp
* 23580 24288: contig of 709 bp in length
* 24289 24388: gap of 100 bp
* 24389 25117: contig of 729 bp in length
* 25118 25217: gap of 100 bp
* 25218 25922: contig of 705 bp in length
* 25923 26022: gap of 100 bp
* 26023 26765: contig of 743 bp in length
* 26766 26865: gap of 100 bp
* 26866 27601: contig of 736 bp in length
* 27602 27701: gap of 100 bp
* 27702 28413: contig of 712 bp in length
* 28414 28513: gap of 100 bp
* 28514 29246: contig of 733 bp in length
* 29247 29346: gap of 100 bp
* 29347 30085: contig of 739 bp in length
* 30086 30185: gap of 100 bp
* 30186 30897: contig of 712 bp in length
* 30898 30997: gap of 100 bp
* 30998 31711: contig of 714 bp in length
* 31712 31811: gap of 100 bp
* 31812 32528: contig of 717 bp in length
* 32529 32628: gap of 100 bp
* 32629 33340: contig of 712 bp in length
* 33341 33440: gap of 100 bp
* 33441 34133: contig of 695 bp in length
* 34136 34235: gap of 100 bp
* 34236 34937: contig of 702 bp in length
* 34938 35037: gap of 100 bp
* 35038 35758: contig of 721 bp in length
* 35759 35858: gap of 100 bp
* 35859 36547: contig of 689 bp in length
* 36548 36647: gap of 100 bp
* 36648 37345: contig of 698 bp in length
* 37346 37445: gap of 100 bp

```

```

* 37446 38186: contig of 741 bp in length
* 38187 38286: gap of 100 bp
* 38287 38996: contig of 710 bp in length
* 38997 39096: gap of 100 bp
* 39097 39787: contig of 691 bp in length
* 39788 39887: gap of 100 bp
* 39888 40608: contig of 721 bp in length
* 40609 40708: gap of 100 bp
* 40709 41434: contig of 726 bp in length
* 41435 41534: gap of 100 bp
* 41535 42253: contig of 719 bp in length
* 42254 42353: gap of 100 bp
* 42354 43056: contig of 703 bp in length
* 43057 43156: gap of 100 bp
* 43157 43866: contig of 710 bp in length
* 43867 43966: gap of 100 bp
* 43967 44675: contig of 709 bp in length
* 44676 44775: gap of 100 bp
* 44776 45483: contig of 708 bp in length
* 45484 45583: gap of 100 bp
* 45584 46311: contig of 728 bp in length
* 46312 46411: gap of 100 bp
* 46412 47129: contig of 718 bp in length
* 47130 47229: gap of 100 bp
* 47230 47941: contig of 712 bp in length
* 47942 48041: gap of 100 bp
* 48042 48766: contig of 725 bp in length
* 48767 48866: gap of 100 bp
* 48867 49592: contig of 726 bp in length
* 49593 49692: gap of 100 bp
* 49693 50404: contig of 712 bp in length
* 50405 50504: gap of 100 bp
* 50505 51202: contig of 698 bp in length
* 51203 51302: gap of 100 bp
* 51303 52008: contig of 706 bp in length
* 52009 52108: gap of 100 bp
* 52109 52810: contig of 702 bp in length
* 52811 52910: gap of 100 bp
* 52911 53642: contig of 732 bp in length
* 53643 53742: gap of 100 bp
* 53743 54459: contig of 717 bp in length
* 54460 54559: gap of 100 bp
* 54560 55267: contig of 708 bp in length
* 55268 55367: gap of 100 bp
* 55368 56077: contig of 710 bp in length
* 56078 56177: gap of 100 bp

```

Query Match 79.2%; Score 19.8; DB 78; Length 64231;  
 Best Local Similarity 91.3%; Pred. No. 20;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 acattgtcgcaagtccacagaatt 24  
 |||||  
 DB 32875 ACATTGCGCAAAATGACACAATT 32853

RESULT 20

AC016743  
 LOCUS AC016743 144355 bp DNA PRI 24-MAR-2001  
 DEFINITION Homo sapiens clone Rp11-475H17, complete sequence.  
 ACCESSION AC016743  
 VERSION AC016743.10 GI:13443262  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 144355)  
 AUTHORS Waterston,R.H.  
 TITLE The sequence of Homo sapiens clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 144355)  
 AUTHORS Waterston,R.H.

TITLE Direct Submission  
JOURNAL Submitted (04-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 3 (bases 1 to 144355)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (24-MAR-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
COMMENT On Mar 24, 2001 this sequence version replaced gi:13270791.  
Center project name: H\_NH0475H17.  
FEATURES  
source  
1..144355  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-475H17"  
BASE COUNT 48459 a 27548 c 25823 g 42525 t  
ORIGIN  
Query Match 79.2%; Score 19.8; DB 87; Length 144355;  
Best Local Similarity 91.3%; Pred. No. 20;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 2 acattgtcgaagtccagaatt 24  
||||||| ||||| ||||| ||||| |||||  
Db 107136 ACATTGTCACAAAGTCACAGAATT 107158  
RESULT 21  
AC019147/c  
LOCUS AC019147 194576 bp DNA HTG 07-JUL-2000  
DEFINITION Homo sapiens chromosome 17 clone RP11-763E3, WORKING DRAFT  
SEQUENCE, 30 uncloned pieces.  
AC019147  
VERSION AC019147.4 GI:8570396  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 194576)  
AUTHORS Waterston, R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 194576)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (30-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
COMMENT On Jun 17, 2000 this sequence version replaced gi:7235343.  
----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
----- Project Information -----  
Center project name: H\_NH0763E03  
----- Summary Statistics -----  
Sequencing vector: M13; 85%  
Sequencing vector: plasmid; 15%  
Chemistry: Dye-primer ET; 85% of reads  
Chemistry: Dye-terminator Big Dye; 15% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 168961 bases at least Q40  
Consensus quality: 176420 bases at least Q30  
Consensus quality: 180507 bases at least Q20  
Insert size: 178000; agarose-fp  
Insert size: 191676; sum-of-contigs  
Quality coverage: 3.44 in Q20 bases; agarose-fp  
Quality coverage: 3.31 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 30 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1  
\* 1065: contig of 1065 bp in length  
\* 1165: gap of unknown length  
\* 1166: contig of 1861 bp in length  
\* 3026: contig of 1861 bp in length  
\* 3027: gap of unknown length  
\* 3127: contig of 1732 bp in length  
\* 4858: gap of unknown length  
\* 4859: gap of unknown length  
\* 4959: contig of 1154 bp in length  
\* 6112: gap of unknown length  
\* 6113: contig of 2379 bp in length  
\* 6213: gap of unknown length  
\* 8591: gap of unknown length  
\* 8592: contig of 3459 bp in length  
\* 12150: gap of unknown length  
\* 12151: contig of 2607 bp in length  
\* 14857: gap of unknown length  
\* 14858: contig of 2694 bp in length  
\* 14958: gap of unknown length  
\* 17651: contig of 4111 bp in length  
\* 17652: gap of unknown length  
\* 21862: contig of 4111 bp in length  
\* 21863: gap of unknown length  
\* 21963: contig of 3418 bp in length  
\* 25381: gap of unknown length  
\* 25481: contig of 3698 bp in length  
\* 29179: gap of unknown length  
\* 29279: contig of 4038 bp in length  
\* 33116: gap of unknown length  
\* 33117: contig of 4596 bp in length  
\* 33417: gap of unknown length  
\* 38012: contig of 4992 bp in length  
\* 38013: gap of unknown length  
\* 43104: contig of 4370 bp in length  
\* 43204: gap of unknown length  
\* 47574: contig of 4370 bp in length  
\* 47575: gap of unknown length  
\* 47675: contig of 4958 bp in length  
\* 52632: gap of unknown length  
\* 52732: contig of 4326 bp in length  
\* 57058: gap of unknown length  
\* 57158: contig of 4863 bp in length  
\* 62021: gap of unknown length  
\* 62022: contig of 5208 bp in length  
\* 62122: gap of unknown length  
\* 67329: contig of 5857 bp in length  
\* 67429: gap of unknown length  
\* 73286: contig of 7223 bp in length  
\* 73386: gap of unknown length  
\* 80609: contig of 7223 bp in length  
\* 80709: gap of unknown length  
\* 80710: contig of 6434 bp in length  
\* 87143: gap of unknown length  
\* 87243: contig of 8791 bp in length  
\* 96034: gap of unknown length  
\* 96134: contig of 8506 bp in length  
\* 104640: gap of unknown length  
\* 104740: contig of 9059 bp in length  
\* 113799: gap of unknown length  
\* 113899: contig of 11385 bp in length  
\* 125284: gap of unknown length  
\* 125385: contig of 12378 bp in length  
\* 137762: gap of unknown length  
\* 137763: contig of 14642 bp in length  
\* 137863: gap of unknown length  
\* 152505: contig of 16388 bp in length  
\* 152605: gap of unknown length  
\* 168993: contig of 25484 bp in length  
\* 169093: gap of unknown length  
\* 169093: contig of 25484 bp in length.  
Location/Qualifiers  
1..194576  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"

FEATURES  
source

```

misc_feature      /clone="RP11-763E3"
1. .1065
/note="assembly_name:Contig5"
misc_feature      1166. .3026
/note="assembly_name:Contig6"
misc_feature      3127. .4858
/note="assembly_name:Contig8"
misc_feature      4959. .6112
/note="assembly_name:Contig9"
misc_feature      6213. .8591
/note="assembly_name:Contig12"
misc_feature      8692. 12150
/note="assembly_name:Contig13"
misc_feature      12251. .14857
/note="assembly_name:Contig14"
misc_feature      14958. .17651
/note="assembly_name:Contig16"
misc_feature      17752. .21862
/note="assembly_name:Contig17"
misc_feature      21963. .25380
/note="assembly_name:Contig18"
misc_feature      25481. .29178
/note="assembly_name:Contig19"
misc_feature      29279. .33316
/note="assembly_name:Contig20"
misc_feature      33417. .38012
/note="assembly_name:Contig21"
misc_feature      38113. .43104
/note="assembly_name:Contig22"
misc_feature      43205. .47574
/note="assembly_name:Contig23"
clone_end:SP6
vector_side:left"
47675. .52632
/note="assembly_name:Contig24"
misc_feature      52733. .57058
/note="assembly_name:Contig25"
misc_feature      57159. .62021
/note="assembly_name:Contig26"
misc_feature      62122. .67329
/note="assembly_name:Contig27"
misc_feature      67430. .73286
/note="assembly_name:Contig28"
misc_feature      73387. .80609
/note="assembly_name:Contig29"
misc_feature      80710. .87143
/note="assembly_name:Contig30"
misc_feature      87244. .96034
/note="assembly_name:Contig31"
misc_feature      96135. .104640
/note="assembly_name:Contig32"
misc_feature      104741. .113799
/note="assembly_name:Contig33"
misc_feature      113900. .125284
/note="assembly_name:Contig34"
misc_feature      125385. .137762
/note="assembly_name:Contig35"
misc_feature      137863. .152504
/note="assembly_name:Contig36"
misc_feature      152605. .168992
/note="assembly_name:Contig37"
misc_feature      169093. .194576
/note="assembly_name:Contig38"
BASE COUNT      60015 a 38425 c 37951 g 55254 t 2931 others
ORIGIN

```

```

Query Match      79.2%; Score 19.8; DB 65; Length 194576;
Best Local Similarity 91.3%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 2 acattgtcgcaagtacagaatt 24
|||||

```

Db 155577 ACATTGTGCAAGTACAGAATT 155555

RESULT 22  
AC012205/c

LOCUS  
DEFINITION  
AC012205 228434 bp DNA HTG 30-MAR-2000  
SEQUENCE. 24 unordered pieces.  
AC012205  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
human.  
AC012205.3 GI:7341726  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE  
ORGANISM

Human sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 228434)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 4, clone RP11-340K9  
Unpublished  
2 (bases 1 to 228434)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baldwin,J., Barna,N., Beckerly,R., Boquslavkiy,L., Boukhalter,B.,  
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,  
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,  
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,  
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE

JOURNAL

COMMENT

Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 30, 2000 this sequence version replaced gi:6479150.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L1745

Center clone name: 340.K.9

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 213999 bases at least Q40

Consensus quality: 221073 bases at least Q30

Consensus quality: 223862 bases at least Q20

Insert size: 225000; agarose-fp

Insert size: 226134; sum-of-contigs

Quality coverage: 5.1 in Q20 bases; agarose-fp

Quality coverage: 5.1 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 24 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1012: contig of 1012 bp in length  
\* 1013 1112: gap of 100 bp  
\* 1113 2281: contig of 1169 bp in length  
\* 2282 2381: gap of 100 bp



```

* 2382 3500: contig of 1119 bp in length
* 3501 3600: gap of 100 bp
* 3601 4671: contig of 1071 bp in length
* 4672 4771: gap of 100 bp
* 4772 5843: contig of 1072 bp in length
* 5844 5943: gap of 100 bp
* 5944 6957: contig of 1054 bp in length
* 6958 7097: gap of 100 bp
* 7098 8534: contig of 1437 bp in length
* 8535 8634: gap of 100 bp
* 8635 9966: contig of 1332 bp in length
* 9967 10066: gap of 100 bp
* 10067 11613: contig of 1547 bp in length
* 11614 11713: gap of 100 bp
* 11714 13991: contig of 2278 bp in length
* 13992 14091: gap of 100 bp
* 14092 15474: contig of 1383 bp in length
* 15475 15574: gap of 100 bp
* 15575 17959: contig of 2385 bp in length
* 17960 18059: gap of 100 bp
* 18060 20530: contig of 2471 bp in length
* 20531 20630: gap of 100 bp
* 20631 25117: contig of 4487 bp in length
* 25118 25217: gap of 100 bp
* 25218 32320: contig of 7103 bp in length
* 32321 32420: gap of 100 bp
* 32421 40992: contig of 8572 bp in length
* 40993 41092: gap of 100 bp
* 41093 49538: contig of 8446 bp in length
* 49539 49638: gap of 100 bp
* 49639 57982: contig of 8344 bp in length
* 57983 58082: gap of 100 bp
* 58083 73752: contig of 15670 bp in length
* 73753 73852: gap of 100 bp
* 73853 90559: contig of 16707 bp in length
* 90560 90659: gap of 100 bp
* 90660 108100: contig of 17441 bp in length
* 108101 108200: gap of 100 bp
* 108201 128050: contig of 19850 bp in length
* 128051 128150: gap of 100 bp
* 128151 172825: contig of 44675 bp in length
* 172826 172925: gap of 100 bp
* 172926 228434: contig of 55509 bp in length.
FEATURES
    source
        Location/Qualifiers
            1..228434
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="4"
                /map="4"
                /clone="RP11-340K9"
                /clone_lib="RPC1-11 Human Male BAC"
            1..1012
                /note="assembly_fragment"
            1113..2281
                /note="assembly_fragment"
            2382..3500
                /note="assembly_fragment"
            3601..4671
                /note="assembly_fragment"
            4772..5843
                /note="assembly_fragment"
            5944..6997
                /note="assembly_fragment"
            7098..8534
                /note="assembly_fragment"
            8635..9966
                /note="assembly_fragment"
            10067..11613
                /note="assembly_fragment"
            11714..13991
                /note="assembly_fragment"
            14092..15474
                /note="assembly_fragment"
            15475..17959
                /note="assembly_fragment"

```

```

misc_feature
    15575..17959
        /note="assembly_fragment"
        clone_end:T7
        vector_side:left"
    18060..20530
        /note="assembly_fragment"
        clone_end:SP6
        vector_side:right"
    20631..25117
        /note="assembly_fragment"
    25218..32320
        /note="assembly_fragment"
    32421..40992
        /note="assembly_fragment"
    41093..49538
        /note="assembly_fragment"
    49639..57982
        /note="assembly_fragment"
    58083..73752
        /note="assembly_fragment"
    73853..90559
        /note="assembly_fragment"
    90660..108100
        /note="assembly_fragment"
    108201..128050
        /note="assembly_fragment"
    128151..172825
        /note="assembly_fragment"
    172926..228434
        /note="assembly_fragment"
BASE COUNT 73186 a 43737 c 41722 g 67487 t 2302 others
ORIGIN

```

```

Query Match 79.2%; Score 19.8; DB 62; Length 228434;
Best Local Similarity 91.3%; Pred. NO. 20;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 2 acattgtcgaagtcacagaatt 24
||||||| ||||| ||||| |||||

```

```

Db 19903 ACATTGTCACAAAGTGACAGAAAT 19881

```

```

RESULT 23
CEL39D10/c CEL39D10 40897 bp DNA INV 02-NOV-1995
LOCUS CEL39D10
DEFINITION Caenorhabditis elegans cosmid C39D10.
ACCESSION U39678
VERSION U39678.1 GI:1049439
KEYWORDS
SOURCE
    Caenorhabditis elegans strain-Bristol N2.
ORGANISM
    Caenorhabditis elegans
    Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
    Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
    1 (bases 1 to 40897)
AUTHORS
    Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
    Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
    Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
    Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
    Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
    Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
    O'Callaghan,M., Parsons,J., Percy,C., Rifkin,L., Roopra,A.,
    Saunders,D., Showkneen,R., Smaildon,N., Smith,A., Sonhammer,E.,
    Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
    Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
    Wilkison-Sproat,J. and Wohlman,P.
    2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
    elegans

```

```

JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 40897)
AUTHORS Minx,P.
TITLE The sequence of C. elegans cosmid C39D10

```







```
Center clone name: RP11-73L6
----- Summary Statistics
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 132601 bases at least Q40
Consensus quality: 140586 bases at least Q30
Consensus quality: 144168 bases at least Q20
Estimated insert size: 144485; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 49434: contig of 49434 bp in length
* 49435 49534: gap of unknown length
* 49535 76020: contig of 26486 bp in length
* 76021 76120: gap of unknown length
* 76121 106128: contig of 30008 bp in length
* 106129 106228: gap of unknown length
* 106229 130157: contig of 23929 bp in length
* 130158 130257: gap of unknown length
* 130258 139669: contig of 9412 bp in length
* 139670 139769: gap of unknown length
* 139770 143465: contig of 5696 bp in length
* 143466 145565: gap of unknown length
* 145566 146585: contig of 1020 bp in length.
*
* Location/Qualifiers
*     source          1. 146585
*     organism="Homo sapiens"
*     db_xref="taxon:9606"
*     /chromosome="3"
*     /clone="RP11-73L6"
*
* BASE COUNT      43675 a 28441 c 28398 g 45447 t 624 others
* ORIGIN
*
* Query Match      76.8%; Score 19.2; DB 72; Length 146585;
* Best Local Similarity 87.5%; Pred. NO. 41;
* Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
*
* QY      2 acattgtcgaagtccacagaatta 25
*          ||||| | | | | | | | | | | | | | | | |
* Db 146028 ACATTGCTCTAGTCACATAATTA 146051
*
* RESULT 26
* AC063937/c
* LOCUS      AC063937      149971 bp      DNA      25-JUL-2000
* DEFINITION Homo sapiens chromosome 3 clone RP11-79A14, WORKING DRAFT SEQUENCE,
*              13 unordered pieces.
* ACCESSION      AC063937
* VERSION      AC063937.8 GI:9438590
* KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
* SOURCE      human.
* ORGANISM      Homo sapiens
* Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
* Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
* 1 (bases 1 to 149971)
* Bodota.B., Bouck.J., Bowie.S., Brooks.A., Buhay.C., Bunac.C.,
* Burkett.C., Burrows.J., Carter.M., Chacko.J., Chen.Z., Cox.C.,
* David.R., Delgado.O., Deshazo.D., Ding.Y., Domah-Rashid.N.,
* Dugan-Rocha.S., Durbin.K.J., Fernandez.C., Ferraguto.D.,
* Forcum-Tansey.J., Frantz.P., Ganesh.R., Gorrell.J.H., Gorrell.L.L.,
```

```

Guevara.W., Harris.K., Hernandez.J., Hodgson.A., Hoques.M.,
Holloway.C., Hosak.H., Jackson.L.E., Jackson.L., Jia.Y., Jones.M.,
Kelly.S., Kondejewski.N., Kong.Y., Kovar.C., Leal.B., Li.Z.,
Lichter.O., Liu.J., Liu.W., Logan.O., Lozano.R.J., Lu.J.,
Lucier.R., Martin.R., Martinez.C., McLeod.M.P., Mel.G., Morgan.M.,
Morris.S., Nash.S., Nelson.A., Nguyen.R., Nguyen.N., Nguyen.S.,
Oswal.G., Parish.R., Paxton.S., Payton.B., Perez.L., Pu.L.L.,
Quiles.M., Reiter.D., Rives.M., Samuel.S., Say.J., Scherer.S.,
Shah.E., Shen.H., Simon.M., Sparks.A., Stamps.A., Sugang.R.,
Tabor.P., Taylor.T., Vasquez.L., Vinson.R., Vo.O., Wabnitz.M.,
Watlington.S., Weinstein.G., Weinstein.I.R., Williamson.A.,
Worley.K., Wren.J., Wrenford.G., Yu.W., Zhou.X., Nelson.D. and
Gibbs.R.
Direct Submission
Unpublished
2 (bases 1 to 149971)
Worley.K.C.
Direct Submission
Submitted (22-APR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 25, 2000 this sequence version replaced gi:8705281.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: RP11-79A14
Center clone name: RP11-79A14
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 130792 bases at least Q40
Consensus quality: 142171 bases at least Q30
Consensus quality: 145357 bases at least Q20
Estimated insert size: 145483; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 3.6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1. 46806: contig of 46806 bp in length
* 46807 46906: gap of unknown length
* 46907 79857: contig of 32951 bp in length
* 79858 79957: gap of unknown length
* 79958 99924: contig of 19967 bp in length
* 99925 100024: gap of unknown length
* 100025 116534: contig of 16510 bp in length
* 116535 116634: gap of unknown length
* 116635 122163: contig of 5529 bp in length
* 122164 122263: gap of unknown length
* 122264 128397: contig of 6133 bp in length
* 128397 128496: gap of unknown length
* 128497 134588: contig of 6092 bp in length
* 134589 134688: gap of unknown length
* 134689 138579: contig of 3891 bp in length
* 138580 138679: gap of unknown length
* 138680 141815: contig of 3136 bp in length
* 141816 141915: gap of unknown length
* 141916 144532: contig of 2617 bp in length
* 144533 144632: gap of unknown length
* 144633 146694: contig of 2062 bp in length
* 146695 146794: gap of unknown length
* 146795 148808: contig of 2014 bp in length
* 148809 148908: gap of unknown length
```

FEATURES	source
BASE COUNT	44656 a 29149 c 28710 g 46204 t 1252 others
ORIGIN	
Query Match	76.8%; Score 19.2; DB 72; Length 149971;
Best Local Similarity	87.5%; Pred. No. 41;
Matches	21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY	2 acattgtcgaagtcacagaatta 25 
Db	24626 ACATTGTCTCTAGTCACATAAATTA 24603
RESULT 27	
AL161796/c	
LOCUS	
DEFINITION	AL161796 157223 bp DNA HTG 10-AUG-2000
DESCRIPTION	Homo sapiens chromosome 1 clone RP11-487023 map q31.1-31.3, *** SEQUENCING IN PROGRESS ***, 15 unordered pieces.
ACCESSION	AL161796.5 GI:9796864
VERSION	HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS	human.
SOURCE	
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 157223) Pavitt.R. Direct Submission Submitted (09-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk requests: clonerequest@sanger.ac.uk On Aug 12, 2000 this sequence version replaced gi:9212870. ----- Genome Center Center: Sanger Centre Center code: SC Web site: http://www.sanger.ac.uk Contact: humquery@sanger.ac.uk ----- Project Information Center project name: BA487023 ----- Summary Statistics Assembly program: XGAP4; version 4.5 Sequencing vector: plasmid; L08752; 100% of reads Chemistry: Dye-terminator Big Dye; 51% of reads Quality: 149752 bases at least Q40 Consensus quality: 152365 bases at least Q30 Consensus quality: 154034 bases at least Q20 Insert size: 155823; sum-of-contigs Insert size: 105263; 71.7% error; agarose-ff Quality coverage: 4.25x in Q20 bases; sum-of-contigs Quality coverage: 6.30x in Q20 bases; agarose-ff ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 15 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * 1 8863: contig of 8863 bp in length * 8864 8963: gap of 100 bp * 8964 12236: contig of 3273 bp in length * 12237 12336: gap of 100 bp * 12337 28375: contig of 16039 bp in length * 28376 28475: gap of 100 bp

```

misc_feature 129435..157223
/note="assembly_fragment:00952"
BASE COUNT 51854 a 26045 c 25566 g 52353 t 1405 others
ORIGIN

Query Match 76.8%; Score 19.2; DB 79; Length 157223;
Best Local Similarity 87.5%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gacattgtcgaagtcacagaatt 24
||||||| ||| |||||
Db 63418 GACATTGTCTCAATGACAGAATT 63395

RESULT 28
AC068365 167292 bp DNA HTG 25-JUN-2000
LOCUS Homo sapiens chromosome 3 clone RP11-764C7 map 3, WORKING DRAFT
DEFINITION SEQUENCE, 31 unordered pieces.
AC068365
AC068365
AC068365.2 GI:8705056
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 167292)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 3, clone RP11-764C7
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 167292)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgaltier,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glade,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,J., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 25, 2000 this sequence version replaced gi:7677736.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L10186
Center clone name: 764_C_7
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 151203 bases at least Q40

```

Consensus quality: 159094 bases at least Q30  
Consensus quality: 161993 bases at least Q20  
Insert size: 170000; agarose-fp  
Insert size: 164292; sum-of-contigs  
Quality coverage: 3.7 in Q20 bases; agarose-fp  
Quality coverage: 3.8 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 31 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1173: contig of 1173 bp in length  
1 1174 1273: gap of 100 bp  
1 1274 2519: contig of 1246 bp in length  
1 2520 2619: gap of 100 bp  
1 2620 4794: contig of 2175 bp in length  
1 4795 4894: gap of 100 bp  
1 4895 6532: contig of 1638 bp in length  
1 6533 6632: gap of 100 bp  
1 6633 7724: contig of 1092 bp in length  
1 7725 7824: gap of 100 bp  
1 7825 9453: contig of 1629 bp in length  
1 9454 9553: gap of 100 bp  
1 9554 10857: contig of 1304 bp in length  
1 10858 10957: gap of 100 bp  
1 10958 12737: contig of 1780 bp in length  
1 12738 12837: gap of 100 bp  
1 12838 15810: contig of 2973 bp in length  
1 15811 15910: gap of 100 bp  
1 15911 18882: contig of 2972 bp in length  
1 18883 18982: gap of 100 bp  
1 18983 21649: contig of 2667 bp in length  
1 21650 21749: gap of 100 bp  
1 21750 24680: contig of 2931 bp in length  
1 24681 24780: gap of 100 bp  
1 24781 28513: contig of 3733 bp in length  
1 28514 28613: gap of 100 bp  
1 28614 31816: contig of 3203 bp in length  
1 31817 31916: gap of 100 bp  
1 31917 35271: contig of 3355 bp in length  
1 35272 35371: gap of 100 bp  
1 35372 38395: contig of 3024 bp in length  
1 38396 38495: gap of 100 bp  
1 38496 41955: contig of 3460 bp in length  
1 41956 42055: gap of 100 bp  
1 42056 45874: contig of 3819 bp in length  
1 45875 45974: gap of 100 bp  
1 45975 50695: contig of 4721 bp in length  
1 50696 50795: gap of 100 bp  
1 50796 55055: contig of 4260 bp in length  
1 55056 55155: gap of 100 bp  
1 55156 60637: contig of 5482 bp in length  
1 60638 60737: gap of 100 bp  
1 60738 65254: contig of 4517 bp in length  
1 65255 65354: gap of 100 bp  
1 65355 70460: contig of 5106 bp in length  
1 70461 70560: gap of 100 bp  
1 70561 78592: contig of 8032 bp in length  
1 78593 78692: gap of 100 bp  
1 78693 86070: contig of 7378 bp in length  
1 86071 86170: gap of 100 bp  
1 86171 93277: contig of 7107 bp in length  
1 93278 93377: gap of 100 bp  
1 93378 101958: contig of 8581 bp in length  
1 101959 102058: gap of 100 bp  
1 102059 113198: contig of 11140 bp in length  
1 113199 113298: gap of 100 bp  
1 113299 126286: contig of 12988 bp in length  
1 126287 126386: gap of 100 bp

\* 126387 140169: contig of 13783 bp in length  
 \* 140170 140269: gap of 100 bp  
 \* 140270 167292: contig of 27023 bp in length.

## FEATURES

## Source

```

1. 167292
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="3"
  /map="3"
  /clone="RP11-764C7"
  /clone_lib="RPC1-11 Human Male BAC"
1. 1173
  /note="assembly_fragment"
1274. .2519
  /note="assembly_fragment"
2620. .4794
  /note="assembly_fragment"
4895. .6532
  /note="assembly_fragment"
6633. .7724
  /note="assembly_fragment"
7825. .9453
  /note="assembly_fragment"
9554. .10857
  /note="assembly_fragment"
10958. .12737
  /note="assembly_fragment"
12838. .15810
  /note="assembly_fragment"
15911. .18882
  /note="assembly_fragment"
18983. .21649
  /note="assembly_fragment"
21750. .24680
  /note="assembly_fragment"
24781. .28513
  /note="assembly_fragment"
28614. .31816
  /note="assembly_fragment"
31917. .35271
  /note="assembly_fragment"
35372. .38395
  /note="assembly_fragment"
  clone_end:T7
  vector_side:right"
38496. .41955
  /note="assembly_fragment"
42056. .45874
  /note="assembly_fragment"
45975. .50695
  /note="assembly_fragment"
50796. .55055
  /note="assembly_fragment"
55156. .60637
  /note="assembly_fragment"
60738. .65254
  /note="assembly_fragment"
65355. .70460
  /note="assembly_fragment"
  clone_end:SP6
  vector_side:left"
70561. .78592
  /note="assembly_fragment"
78693. .86070
  /note="assembly_fragment"
86171. .93277
  /note="assembly_fragment"
93378. .101958
  /note="assembly_fragment"
102059. .113198
  /note="assembly_fragment"

```

Query Match

76.8%; Score 19.2; DB 73; Length 167292;

Best Local Similarity 87.5%; Pred. No. 41;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 2 acatttcgcaagtacagaaatta 25  
 ||||| ||||| ||||| |||||  
 Db 134143 ACATTGTCCTCAAGTACATAATTA 134166

## RESULT 29

AC040890 173532 bp DNA HTG 11-APR-2000  
 Homo sapiens chromosome 2 clone RP11-362I13 map 2, WORKING DRAFT  
 SEQUENCE, 21 unordered pieces.  
 AC040890  
 AC040890.1 GI:7533996  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 173532)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
 Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,  
 Campoliano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,  
 Collymore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L6456  
 Center clone name: 362.I.13  
 ----- Summary Statistics  
 Sequencing vector: M13; M7815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 164477 bases at least Q40  
 Consensus quality: 168698 bases at least Q30  
 Consensus quality: 170320 bases at least Q20  
 Insert size: 171532; sum-of-contigs  
 Quality coverage: 4.8 in Q20 bases; sum-of-contigs

TITLE  
JOURNAL

## COMMENT

\*\*\*\*\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 21 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is





Chemistry: Dye-primer ET; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 178101 bases at least Q40  
 Consensus quality: 178714 bases at least Q30  
 Consensus quality: 178984 bases at least Q20  
 Insert size: 188000; agarose-fp  
 Quality coverage: 6.93 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 5 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 74370: contig of 74370 bp in length  
 74371 74470: gap of unknown length  
 74471 85434: contig of 10964 bp in length  
 85435 85534: gap of unknown length  
 85535 104284: contig of 18750 bp in length  
 104285 104384: gap of unknown length  
 104385 122620: contig of 18236 bp in length  
 122621 122721: gap of unknown length  
 122721 180721: contig of 58001 bp in length.

## FEATURES

source  
 1..180721  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="3"  
 /clone="RP11-268022"  
 misc\_feature  
 1..74370  
 /note="assembly\_name:Contig10"  
 74471..85434  
 /note="assembly\_name:Contig6"  
 misc\_feature  
 85535..104284  
 /note="assembly\_name:Contig7"  
 clone\_end:SP6  
 vector\_side:right  
 104385..122620  
 /note="assembly\_name:Contig8"  
 122721..180721  
 /note="assembly\_name:Contig9"  
 BASE COUNT 57067 a 32188 c 32085 g 58975 t 402 others  
 ORIGIN

Query Match 76.8%; Score 19.2; DB 74; Length 180721;  
 Best Local Similarity 87.5%; Pred. No. 41;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 acattgtccagtcacagaatta 25  
 ||||| ||||| ||||| |||||  
 Db 38357 ACATTGTCCAGTTACATAATTA 38380

RESULT 31  
 AC026892  
 LOCUS  
 DEFINITION Homo sapiens chromosome 1 clone RP11-286E4, WORKING DRAFT SEQUENCE,  
 3 unordered pieces.  
 ACCESSION AC026892  
 VERSION AC026892.13 GI:12313762  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 190624)  
 AUTHORS Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,

Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,  
 Mao,J., Lam,B., Marathe,R., Miranda,M., Morehouse,A.J., Nguyen,M.,  
 Oelner,P., Palm,C.J., Ramirez,D., Southwick,A.M., Wilhelm,J.,  
 Yu,S. and Davis,R.W.  
 Unpublished  
 2 (bases 1 to 190624)  
 Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,  
 Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,  
 Mao,J., Marathe,R., Morehouse,A.J., Oelner,P., Palm,C.J.,  
 Ramirez,D., Wilhelm,J., Yu,S. and Davis,R.W.  
 Direct Submission  
 Submitted (25-MAR-2000) DNA Sequencing and Technology Center,  
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
 USA

## TITLE

## JOURNAL

## COMMENT

On Jan 19, 2001 this sequence version replaced gi:12203672.  
 ----- Genome Center  
 Center: Stanford DNA Sequencing and Technology Development  
 Center

Center code: SDSTDC

Web site: <http://sequence-www.stanford.edu/group/human/>

Contact: [hum-info@sequence.stanford.edu](mailto:hum-info@sequence.stanford.edu)

----- Project Information

Center project name: 868

Center clone name: RP11-286E4

----- Summary Statistics

Sequencing Vector: M13mp18; X02513

Chemistry: Dye-primer; 0% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 186972 bases at least Q40

Consensus quality: 187490 bases at least Q30

Consensus quality: 187923 bases at least Q20

Insert size: 183887; agarose-fp

Insert size: 190424; sum-of-contigs

Quality coverage: 9.6x in Q20 bases; agarose-fp

Quality coverage: 9.3x in Q20 bases; sum-of-contigs.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 3 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence.

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 15389: contig of 15389 bp in length

\* 15390 15489: gap of unknown length

\* 15490 42164: contig of 26675 bp in length

\* 42165 42264: gap of unknown length

\* 42265 190624: contig of 148360 bp in length.

## FEATURES

## source

1..190624  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RP11-286E4"  
 /clone\_lib="RPC1 human BAC library 11"

## misc\_feature

1..15389  
 /note="assembly\_name:Contig11"

## misc\_feature

15490..42164  
 /note="assembly\_name:Contig12"

## misc\_feature

42265..190624  
 /note="assembly\_name:Contig13"

## BASE COUNT

64094 a 31639 c 31984 g 62704 t 203 others

## ORIGIN

Query Match 76.8%; Score 19.2; DB 70; Length 190624;  
 Best Local Similarity 87.5%; Pred. No. 41;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 gacattgtccagtcacagaatt 24



Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holliday, C., Hollins, B., Homsi, F., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratoch, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Loulseghe, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Meier, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neale, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, F., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Wleczky, R., Wooden, S., Worley, K., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 197719)  
Worley, K.C.

Submitted (15-JUN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Apr 7, 2001 this sequence version replaced gi:13540444.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: HBPY  
Center clone name: RP11-519L1  
----- Summary Statistics  
Sequencing vector: Plasmid; M77789  
Chemistry: Dye-terminator Big Dye; 40% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 196583 bases at least Q40  
Consensus quality: 196777 bases at least Q30  
Consensus quality: 196869 bases at least Q20  
Estimated insert size: 196908; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-gel estimation  
Quality coverage: 8.4x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 142569: contig of 142569 bp in length

\* 142570 142669: gap of unknown length  
\* 142670 197719: contig of 55050 bp in length.

FEATURES  
Source  
1..197719  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clones="RP11-519L1"

BASE COUNT 65559 a 35026 c 34991 g 62043 t 100 others

ORIGIN

Query Match 76.8% Score 19.2; DB 74; Length 197719;  
Best Local Similarity 87.5% Pred. No. 41;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 acattgtgcgaagtcacagaatta 25  
||||||| ||||| ||||| |||||  
Db 72770 ACATTGTCGCAAGTTACATAATTA 72747

RESULT 34  
AC068781  
LOCUS  
DEFINITION Homo sapiens 12 BAC RP11-457K10 (Roswell Park Cancer Institute Human BAC Library) complete sequence.  
AC068781  
AC068781.17 GI:120000439  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 199229)  
Muzny, D.M., Adams, C.C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amaral, H.C., Are, J.R., Banks, T., Barbara, J., Benton, J., Bimaye, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holliday, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Ioshikhes, I., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratoch, J., Kureshi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C., Loulseghe, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Marondel, I., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Merscher, S., Metzger, M., Miller, A., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Montgomery, K.T., Morgan, M., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, F., Telford, B., Thomas, N., Thomas, R., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Wleczky, R., Wooden, S., Worley, K., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., and Gibbs, R.

Wlarczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Kuchelapati, R. and Gibbs, R.

# TITLE JOURNAL

REFERENCE  
2 (bases 1 to 199229)

# AUTHORS

Worley, K.C.

# TITLE JOURNAL

REFERENCE  
3 (bases 1 to 199229)

# AUTHORS

Worley, K.C.

# TITLE JOURNAL

REFERENCE  
3 (bases 1 to 199229)

# AUTHORS

Worley, K.C.

# TITLE JOURNAL

REFERENCE  
3 (bases 1 to 199229)

# AUTHORS

Worley, K.C.

# TITLE JOURNAL

REFERENCE  
3 (bases 1 to 199229)

# AUTHORS

Worley, K.C.

# TITLE JOURNAL

REFERENCE  
3 (bases 1 to 199229)

# AUTHORS

Worley, K.C.

# TITLE JOURNAL

REFERENCE  
3 (bases 1 to 199229)

# AUTHORS

Worley, K.C.

# TITLE JOURNAL

REFERENCE  
3 (bases 1 to 199229)

# AUTHORS

Worley, K.C.

# TITLE JOURNAL

REFERENCE  
3 (bases 1 to 199229)

# AUTHORS

Worley, K.C.

# TITLE JOURNAL

REFERENCE  
3 (bases 1 to 199229)

# AUTHORS

Worley, K.C.

# TITLE JOURNAL

REFERENCE  
3 (bases 1 to 199229)

# AUTHORS

Worley, K.C.

# TITLE JOURNAL

REFERENCE  
3 (bases 1 to 199229)

# AUTHORS

Worley, K.C.

# TITLE JOURNAL

REFERENCE  
3 (bases 1 to 199229)

# AUTHORS

Worley, K.C.

# TITLE JOURNAL

REFERENCE  
3 (bases 1 to 199229)

# AUTHORS

Worley, K.C.

# TITLE JOURNAL

REFERENCE  
3 (bases 1 to 199229)

# AUTHORS

Worley, K.C.

# TITLE JOURNAL

REFERENCE  
3 (bases 1 to 199229)

# AUTHORS

Worley, K.C.

# TITLE JOURNAL

REFERENCE  
3 (bases 1 to 199229)

# AUTHORS

Worley, K.C.

# TITLE JOURNAL

REFERENCE  
3 (bases 1 to 199229)

# AUTHORS

Worley, K.C.

11788 tagactatca(n)tactggagac  
15663 ctgctgggt(n)ttcatccctt  
15952 aactctggag(n)cagtatgttc  
41053 aaacacacac(n)cacacaata  
41135 ttggtatgg(n)cttgcctcta  
41140 tatgnettg(c)ctctatatic  
41191 caaaitaaaa(n)gtgagttccc  
41275 attaatltag(n)gatgtatgt  
41315 aaactctca(n)tgaccagtt  
44792 gcatagttta(n)cattatagtg  
45080 ttggttacag(n)tagttaagaa  
45092 agttaagac(n)ccaggttttc  
46447 cacatgtgca(n)aattgtcagg  
54259 aaccgatct(n)tagtnaaat  
54264 gatctntagt(n)aaatcccaa  
54513 atgcttccat(n)naatcggag  
54514 tnnaaatgga(n)gaagtttct  
54522 ttccctccac(n)ncagccccc  
61194 tctctccacn(n)cagccccc  
61195 tggccacata(n)tggtttttt  
61384 aaagaagag(n)cagataata  
62062 atcaaggtgg(n)caagtaata  
70296 ccactctgta(n)tcngaagtt  
83692 tctgtganc(n)gaagttggc  
83689 ttggtgcttc(n)ccaaatnng  
83707 ttcccaaat(n)ngantctgt  
83714 tcnccaaat(n)ngantctgt  
83716 cnccaaatn(n)gantctgtg  
83719 caaatnnga(n)tcgttggg  
83728 antctgttg(n)gcttgnct  
83735 ttgngccttg(n)tcctgnac  
83742 ttgntcttg(n)cccccagac  
83759 agactctaga(n)aanngaaa  
83760 cctctagana(n)ngngaaaa  
83762 tctaganaa(n)ngngaaaa  
83763 tctaganaa(n)ngngaaaa  
83765 taanaaang(n)gaataatnc  
83770 aannngaaa(a)atnctnnt  
83772 nngngaaaaa(t)tnctnnttt  
83774 gngaaaaat(n)ctnntttta  
83776 gaaaaatnc(t)nttttann  
83777 aaaaatnct(n)tttttann  
83778 aaaaatnct(n)tttttann  
83782 tntctnntt(t)taannctnc  
83785 ctnntttta(n)nnctnctat  
83786 tntttttan(n)nnctnctat  
83787 nnttttann(n)ntnctnctat  
83788 nnttttann(n)ntnctnctat  
83789 ttttannnn(c)tnctnctat  
83790 ttttannnn(c)tnctnctat  
83791 tttannnct(n)ctnctnctat  
83794 annnctnca(n)tttttnttt  
83798 nctnctnctat(t)gtnttttt  
83802 caattttgt(n)ttttttttt  
83803 anttttgt(n)ttttttttt  
83813 tttttttt(t)tnngcccca  
83815 tttttttt(n)ngcccca  
83816 tttttttt(n)ngcccca  
83817 tttttttt(n)gccccaaaa  
83818 tttttttt(n)gccccaaaa  
83819 tttttttnn(g)ccccaaaa  
83826 ngccccaaa(a)ananaaaa  
83828 gccccaaaa(n)ananaaant  
83830 cccaaaaa(n)aaaaaantnt  
83831 caaaaaan(a)aaaaantnt  
83836 ananaaaa(n)ntntaantn  
83837 ananaaaa(n)ntntaantn  
83839 anaaaaant(n)taantntnt  
83840 naaaaaant(n)taantntnt  
83842 aaaaaant(n)taantntnt  
83844 aannntaa(t)ntntntnng  
83845 aannntaa(n)ntntntnng

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:  
STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.  
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.  
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3369-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----  
Contig length: 199229  
Phrap values in estimate: 198842  
Average error rate (BCM-Phrap estimate): 3.80907e-05  
Fraction of Phrap values less than 40 : 0.0105611  
Number of consensus changing edits: 263  
Number of N's in consensus : 0

----- Summary Statistics -----  
Position Original+Context Edited+Context  
6278 gaacaaatcat(n)caaacatcat  
6317 ttgacagtga(n)ngggactcta  
6318 tagcagtga(n)ngggactcta  
6319 tagcagtga(n)ngggactcta  
6322 actctatctc(n)gcaaatgatg  
6361 tcttggtgac(n)ctgggactac

11788 tagactatca(n)tactggagac  
15663 ctgctgggt(n)ttcatccctt  
15952 aactctggag(n)cagtatgttc  
41053 aaacacacac(n)cacacaata  
41135 ttggtatgg(n)cttgcctcta  
41140 tatgnettg(c)ctctatatic  
41191 caaaitaaaa(n)gtgagttccc  
41275 attaatltag(n)gatgtatgt  
41315 aaactctca(n)tgaccagtt  
44792 gcatagttta(n)cattatagtg  
45080 ttggttacag(n)tagttaagaa  
45092 agttaagac(n)ccaggttttc  
46447 cacatgtgca(n)aattgtcagg  
54259 aaccgatct(n)tagtnaaat  
54264 gatctntagt(n)aaatcccaa  
54513 atgcttccat(n)naatcggag  
54514 tnnaaatgga(n)gaagtttct  
54522 ttccctccac(n)ncagccccc  
61194 tctctccacn(n)cagccccc  
61195 tggccacata(n)tggtttttt  
61384 aaagaagag(n)cagataata  
62062 atcaaggtgg(n)caagtaata  
70296 ccactctgta(n)tcngaagtt  
83692 tctgtganc(n)gaagttggc  
83689 ttggtgcttc(n)ccaaatnng  
83707 ttcccaaat(n)ngantctgt  
83714 tcnccaaat(n)ngantctgt  
83716 cnccaaatn(n)gantctgtg  
83719 caaatnnga(n)tcgttggg  
83728 antctgttg(n)gcttgnct  
83735 ttgngccttg(n)tcctgnac  
83742 ttgntcttg(n)cccccagac  
83759 agactctaga(n)aanngaaa  
83760 cctctagana(n)ngngaaaa  
83762 tctaganaa(n)ngngaaaa  
83763 tctaganaa(n)ngngaaaa  
83765 taanaaang(n)gaataatnc  
83770 aannngaaa(a)atnctnnt  
83772 nngngaaaaa(t)tnctnnttt  
83774 gngaaaaat(n)ctnntttta  
83776 gaaaaatnc(t)nttttann  
83777 aaaaatnct(n)tttttann  
83778 aaaaatnct(n)tttttann  
83782 tntctnntt(t)taannctnc  
83785 ctnntttta(n)nnctnctat  
83786 tntttttan(n)nnctnctat  
83787 nnttttann(n)ntnctnctat  
83788 nnttttann(n)ntnctnctat  
83789 ttttannnn(c)tnctnctat  
83790 ttttannnn(c)tnctnctat  
83791 tttannnct(n)ctnctnctat  
83794 annnctnca(n)tttttnttt  
83798 nctnctnctat(t)gtnttttt  
83802 caattttgt(n)ttttttttt  
83803 anttttgt(n)ttttttttt  
83813 tttttttt(t)tnngcccca  
83815 tttttttt(n)ngcccca  
83816 tttttttt(n)ngcccca  
83817 tttttttt(n)gccccaaaa  
83818 tttttttt(n)gccccaaaa  
83819 tttttttnn(g)ccccaaaa  
83826 ngccccaaa(a)ananaaaa  
83828 gccccaaaa(n)ananaaant  
83830 cccaaaaa(n)aaaaaantnt  
83831 caaaaaan(a)aaaaantnt  
83836 ananaaaa(n)ntntaantn  
83837 ananaaaa(n)ntntaantn  
83839 anaaaaant(n)taantntnt  
83840 naaaaaant(n)taantntnt  
83842 aaaaaant(n)taantntnt  
83844 aannntaa(t)ntntntnng  
83845 aannntaa(n)ntntntnng

```

83846      nntnttaatn(n)ntntnnngg      tgcttaataa(c)tagttcagggt
83848      nttaatatnt(n)ttnnngnggt      ttaataatct(a)gttcagggtgt
83849      nttaatatnt(n)ttnnngnggtg      cattaatact(a)gttcagggtgtg
83851      taatntnttt(n)nnngnggtgtg      ttaatacagt(t)cagggtgtg

Query Match      76.8%; Score 19.2; DB 88; Length 199229;
Best Local Similarity 87.5%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2      acattgtcgaagtcacagaatta 25
||||||| ||||| ||||| ||||| |||||
Db 193760 ACATTGTCCCAAGTTACATAATTA 193783

RESULT 35
AC025126      232409 bp      DNA      HTG      15-AUG-2000
LOCUS      Homo sapiens chromosome 18 clone RP11-183N5 map 18; *** SEQUENCING
DEFINITION      IN PROGRESS ***, 92 unordered pieces.
ACCESSION      AC025126
VERSION      AC025126.4 GI:9802846
KEYWORDS      HTG; HTGS_PHASE1.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 232409)
AUTHORS      Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE      Homo sapiens chromosome 18, clone RP11-183N5
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 232409)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castie,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marguis,N.,
McCarthy,M., McEwan,P., McCurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivari,T.M., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
TITLE      Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome
JOURNAL      Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT      On Aug 15, 2000 this sequence version replaced gi:8389580.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WITB
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7774
Center clone name: 183_N_5
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 92 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 315: contig of 315 bp in length

316 415: gap of 100 bp

416 1546: contig of 1131 bp in length

1547 1646: gap of 100 bp

1647 2714: contig of 1068 bp in length

2715 2814: gap of 100 bp

2815 3928: contig of 1114 bp in length

3929 4028: gap of 100 bp

4029 5262: contig of 1234 bp in length

5263 5362: gap of 100 bp

5363 6777: contig of 1415 bp in length

6778 6877: gap of 100 bp

6878 8003: contig of 1126 bp in length

8004 8103: gap of 100 bp

8104 9142: contig of 1039 bp in length

9143 9242: gap of 100 bp

9243 10671: contig of 1429 bp in length

10672 10771: gap of 100 bp

10772 11801: contig of 1030 bp in length

11802 11901: gap of 100 bp

11902 13290: contig of 1389 bp in length

13291 13390: gap of 100 bp

13391 14465: contig of 1075 bp in length

14466 14565: gap of 100 bp

14566 16007: contig of 1442 bp in length

16008 16107: gap of 100 bp

16108 17424: contig of 1317 bp in length

17425 17524: gap of 100 bp

17525 18567: contig of 1043 bp in length

18568 18667: gap of 100 bp

18668 20193: contig of 1526 bp in length

20194 20293: gap of 100 bp

20294 21489: contig of 1196 bp in length

21490 21589: gap of 100 bp

21590 22734: contig of 1145 bp in length

22735 22834: gap of 100 bp

22835 23858: contig of 1024 bp in length

23859 23958: gap of 100 bp

23960 25029: contig of 1071 bp in length

25030 25129: gap of 100 bp

25130 26400: contig of 1271 bp in length

26401 26500: gap of 100 bp

26501 27505: contig of 1005 bp in length

27506 27605: gap of 100 bp

27606 29263: contig of 1658 bp in length

29264 29363: gap of 100 bp

29364 30538: contig of 1175 bp in length

30539 30638: gap of 100 bp

30639 32114: contig of 1476 bp in length

32115 32214: gap of 100 bp

32215 34074: contig of 1860 bp in length

34075 34174: gap of 100 bp

34175 35360: contig of 1186 bp in length

35361 35460: gap of 100 bp

35461 36996: contig of 1536 bp in length

36997 37096: gap of 100 bp

37097 38668: contig of 1572 bp in length

38669 38768: gap of 100 bp

38769 40739: contig of 1971 bp in length

40740 40839: gap of 100 bp

40840 42472: contig of 1633 bp in length

42473 42572: gap of 100 bp

42573 44189: contig of 1617 bp in length

44190 44289: gap of 100 bp

44290 45726: contig of 1437 bp in length

45727 45826: gap of 100 bp

45827 47289: contig of 1463 bp in length

47290 47389: gap of 100 bp

47390 49166: contig of 1777 bp in length

49167 49266: gap of 100 bp

```

* 49267 51468: contig of 2202 bp in length
* 51469 51568: gap of 100 bp
* 51569 52883: contig of 1315 bp in length
* 52884 52983: gap of 100 bp
* 52984 54680: contig of 1697 bp in length
* 54681 54780: gap of 100 bp
* 54781 56584: contig of 1804 bp in length
* 56585 56684: gap of 100 bp
* 56685 58027: contig of 1343 bp in length
* 58028 58127: gap of 100 bp
* 58128 59566: contig of 1439 bp in length
* 59567 59666: gap of 100 bp
* 59667 61605: contig of 1939 bp in length
* 61606 61705: gap of 100 bp
* 61706 62868: contig of 1163 bp in length
* 62869 62968: gap of 100 bp
* 62969 64543: contig of 1575 bp in length
* 64544 64643: gap of 100 bp
* 64644 67120: contig of 2477 bp in length
* 67121 67220: gap of 100 bp
* 67221 68814: contig of 1594 bp in length
* 68815 68914: gap of 100 bp
* 68915 70497: contig of 1583 bp in length
* 70498 70597: gap of 100 bp
* 70598 72725: contig of 2128 bp in length
* 72726 72825: gap of 100 bp
* 72826 74773: contig of 1948 bp in length
* 74774 74873: gap of 100 bp
* 74874 76723: contig of 1850 bp in length
* 76724 76823: gap of 100 bp
* 76824 85187: contig of 8364 bp in length
* 85188 85287: gap of 100 bp
* 85288 88050: contig of 2763 bp in length
* 88051 88150: gap of 100 bp
* 88151 89351: contig of 1201 bp in length
* 89352 89451: gap of 100 bp
* 89452 91285: contig of 1834 bp in length
* 91286 91385: gap of 100 bp
* 91386 94258: contig of 2873 bp in length
* 94259 94358: gap of 100 bp
* 94359 96730: contig of 2372 bp in length
* 96731 96830: gap of 100 bp
* 96831 100362: contig of 3532 bp in length
* 100363 100452: gap of 100 bp
* 100463 102354: contig of 1892 bp in length
* 102355 102454: gap of 100 bp
* 102455 104798: contig of 2344 bp in length
* 104799 104898: gap of 100 bp
* 104899 106543: contig of 1645 bp in length
* 106544 106643: gap of 100 bp
* 106644 108168: contig of 1525 bp in length
* 108169 108268: gap of 100 bp
* 108269 110965: contig of 2697 bp in length
* 110966 111065: gap of 100 bp
* 111066 112766: contig of 1701 bp in length
* 112767 112866: gap of 100 bp
* 112867 115404: contig of 2538 bp in length
* 115405 115504: gap of 100 bp
* 115505 117988: contig of 2484 bp in length
* 117989 118088: gap of 100 bp
* 118089 121235: contig of 3147 bp in length
* 121236 121335: gap of 100 bp
* 121336 123563: contig of 2228 bp in length
* 123564 123663: gap of 100 bp
* 123664 125656: contig of 1993 bp in length
* 125657 125756: gap of 100 bp
* 125757 128559: contig of 2803 bp in length
* 128560 128659: gap of 100 bp
* 128660 131968: contig of 3309 bp in length
* 131969 132068: gap of 100 bp

```

Query Match 76.8%; Score 19.2; DB 69; Length 232409;  
 Best Local Similarity 87.5%; Pred. No. 41;

```

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 acattgtcgcaagtcacagattta 25
||||| ||| |||||||
Db 58572 ACATTGTCCCAATAACAGATTTA 58595

RESULT 36
HS2A2A/c 32671 bp DNA PRI 23-NOV-1999
LOCUS Human DNA sequence from PAC 2A2 on chromosome X.
DEFINITION Z84815
ACCESSION Z84815
VERSION 284815.1 GI:1834462
KEYWORDS X.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 32671)
Direct Submission
Submitted (04-FEB-1997) E-mail enquires: humquery@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
de Jong P.J., enquires: http://bacpac.med.buffalo.edu/ IMPORTANT:
This sequence is not the entire insert of clone 2A2. It may be
shorter because we only sequence overlapping sections once, or
longer because we arrange for a small overlap between neighbouring
submissions.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The sequence from clone 2A2 has been finished in more than one
contig. This sequence (2A2A) is separated from the following one
(2A2B) by a gap of 350 bp sized by PCR.
The true left end of clone 2A2 is at 1 in this sequence. 2A2 is
from the human PAC library described in Ioannou A.P. et al Nature
Genet 6, 84-89.

FEATURES
source
1..32671
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone_lib="RPCI-1"
/clone="RPI-2A2"
659..1020
/note="LI repeat: matches 4769. 4409 of consensus"
repeat_region
1005..1354
/note="LI repeat: matches 1. 391 of consensus"
repeat_region
2554..2841
/note="LI repeat: matches 4423. 4336 of consensus"
repeat_region
2641..2759
/note="LI repeat: matches 4785. 4904 of consensus"
repeat_region
2751..3498
/note="LI repeat: matches 147. 909 of consensus"
repeat_region
5034..5084
/note="LI repeat: matches 401. 351 of consensus"
repeat_region
5079..5387
/note="LI repeat: matches 1. 309 of consensus"
repeat_region
5079..5635
/note="LI repeat: matches 1. 546 of consensus"
repeat_region
5642..5979
/note="LI repeat: matches 358. 1 of consensus"
repeat_region
6162..6332
/note="LI repeat: matches 4586. 4756 of consensus"
repeat_region
6503..7375
/note="LI repeat: matches 902. 1 of consensus"
repeat_region
7239..11641
/note="LI repeat: matches 5390. 966 of consensus"
repeat_region
13509..13800

```

repeat\_region /note="AluX repeat: matches 1. .293 of consensus"  
 13803. .14023  
 repeat\_region /note="L1ME1 repeat: matches 860. .655 of consensus"  
 14047. .14383  
 repeat\_region /note="L1ME2 repeat: matches 672. .334 of consensus"  
 14385. .14679  
 repeat\_region /note="AluX repeat: matches 1. .302 of consensus"  
 14815. .14862  
 repeat\_region /note="AluJ repeat: matches 302. .254 of consensus;  
 incomplete repeat"  
 14884. .15035  
 repeat\_region /note="L1MB5 repeat: matches 216. .63 of consensus"  
 15227. .15574  
 repeat\_region /note="THE1B repeat: matches 361. .5 of consensus"  
 15988. .16442  
 repeat\_region /note="L1MA9 repeat: matches 1017. .563 of consensus"  
 16544. .16854  
 repeat\_region /note="L1MA2 repeat: matches 1. .315 of consensus"  
 17046. .17342  
 repeat\_region /note="AluSc repeat: matches 297. .1 of consensus"  
 17963. .18091  
 repeat\_region /note="L1MB8 repeat: matches 911. .778 of consensus"  
 18106. .18563  
 repeat\_region /note="MLT2CB repeat: matches 1. .461 of consensus"  
 18619. .18663  
 repeat\_region /note="MLT2CB repeat: matches 457. .501 of consensus"  
 18665. .20109  
 repeat\_region /note="MLT2-internal repeat: matches 2. .1389 of consensus"  
 20135. .20340  
 repeat\_region /note="L1MA7 repeat: matches 236. .442 of consensus"  
 20331. .20720  
 repeat\_region /note="L1HS repeat: matches 395. .1 of consensus"  
 20575. .23701  
 repeat\_region /note="L1 repeat: matches 5390. .2254 of consensus"  
 23707. .24060  
 repeat\_region /note="THE1C repeat: matches 360. .1 of consensus"  
 24075. .24914  
 repeat\_region /note="L1 repeat: matches 2242. .1403 of consensus"  
 24718. .25249  
 repeat\_region /note="MER25 repeat: matches 2101. .1583 of consensus"  
 26154. .26302  
 repeat\_region /note="L1 repeat: matches 4200. .4059 of consensus"  
 26302. .26366  
 repeat\_region /note="THE1C repeat: matches 371. .305 of consensus"  
 26334. .26842  
 repeat\_region /note="L1MA2 repeat: matches 532. .1055 of consensus"  
 27345. .27453  
 repeat\_region /note="AluJb repeat: matches 1. .107 of consensus;  
 incomplete repeat"  
 27461. .27812  
 repeat\_region /note="THE1B repeat: matches 364. .1 of consensus"  
 27813. .29193  
 repeat\_region /note="THE1B-INTERNAL repeat: matches 1580. .198 of  
 consensus"  
 29194. .29488  
 repeat\_region /note="AluSg repeat: matches 300. .2 of consensus"  
 29489. .29692  
 repeat\_region /note="THE1B-INTERNAL repeat: matches 206. .1 of consensus"  
 29693. .30040  
 repeat\_region /note="THE1B repeat: matches 364. .4 of consensus"  
 30049. .30210  
 repeat\_region /note="AluJo repeat: matches 114. .282 of consensus;  
 incomplete repeat"  
 30966. .32635  
 repeat\_region /note="L1 repeat: matches 3679. .5382 of consensus"  
 32498. .32671  
 repeat\_region /note="L1MC2 repeat: matches 3. .178 of consensus"  
 32552 a 5758 c 6033 g 11328 t

Query Match 75.2%; Score 18.8; DB 92; Length 32671;  
 Best Local Similarity 90.9%; Pred. No. 67;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 2 acattgtcgaagtacagaat 23  
 ||||| ||||| |||||  
 Db 26394 ACATTGTGCAAGTGACAGAAT 26373  
 RESULT 37  
 AC079391/c DNA HTG 01-SEP-2000  
 LOCUS Homo sapiens chromosome UNK clone CTD-2243M11, WORKING DRAFT  
 DEFINITION SEQUENCE, 8 unordered pieces.  
 AC079391  
 ACCESSION AC079391.1 GI:9954808  
 VERSION HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS human.  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 108509)  
 AUTHORS Waterston, R.H.  
 TITLE The sequence of Homo sapiens clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 108509)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-AUG-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA

## COMMENT

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 ----- Project Information -----  
 Center project name: H\_MS2243M11  
 ----- Summary Statistics -----  
 Sequencing vector: M13; 100%  
 Sequencing method: plasmid; 0%  
 Chemistry: Dye-primer ET; 100% of reads  
 Assembly program: Phrap; version 0.950319  
 Consensus quality: 105383 bases at least Q40  
 Consensus quality: 106385 bases at least Q30  
 Consensus quality: 106812 bases at least Q20  
 Insert size: 113000; agarose-fp  
 Insert size: 107809; sum-of-contigs  
 Quality coverage: 6.15 in Q20 bases; agarose-fp  
 Quality coverage: 6.49 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 1972: contig of 1972 bp in length  
 \* 1973 2072: gap of unknown length  
 \* 2073 6159: contig of 4087 bp in length  
 \* 6160 6259: gap of unknown length  
 \* 6260 14090: contig of 7831 bp in length  
 \* 14091 14190: gap of unknown length  
 \* 14191 24964: contig of 10774 bp in length  
 \* 24965 25064: gap of unknown length  
 \* 25065 36317: contig of 11253 bp in length  
 \* 36318 36417: gap of unknown length  
 \* 36418 52199: contig of 15782 bp in length  
 \* 52199 52200: gap of unknown length  
 \* 52200 74123: contig of 21824 bp in length  
 \* 74124 108509: contig of 34286 bp in length.  
 \* 74224









phrap Value Range

**CLONE LENGTH:** This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the

```

Version: 1.01 qxfo.
Location/Qualifiers
1. .105207
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12p12-27.2-31.7"
/clone="RP11-16J10"
repeat_region
18...60
/rpt_family="TA)n"
63...190
/standard_name="G41297"
/db_xref="dbSTS:61467"
complement(494...724)
/rpt_family="MIR"
repeat_region
complement(1141...2037)
/rpt_family="L1MC/D"
2118...2527
/rpt_family="LTR52"
3450...3542
/rpt_family="T-rich"
complement(4520...4976)
/rpt_family="HAL1"
repeat_region
complement(5590...5769)
/rpt_family="LTR50"
repeat_region
complement(6173...6317)
/rpt_family="LTR41"
repeat_region
complement(7296...7605)
/rpt_family="AluSp"
complement(8517...8632)
/rpt_family="MLT1A1"
repeat_region
complement(10161...10450)
/rpt_family="AluY"
11042...11157
/rpt_family="L2"
repeat_region
complement(12297...12472)
/rpt_family="MIR"
12499...12537
/rpt_family="L2"
repeat_region
complement(12558...12933)
/rpt_family="MLT1A1"
13472...13730
/standard_name="G41298"
/db_xref="dbSTS:61468"
complement(13981...14015)
/rpt_family="L2"
14109...14172
/rpt_family="MIR"
complement(14173...14808)
/rpt_family="L1MA5"
14809...14822
/rpt_family="MIR"
complement(14919...15071)
/rpt_family="L2"
repeat_region
complement(15133...15254)
/rpt_family="MIR"
repeat_region
complement(15938...16049)
/rpt_family="L1MA3"
18921...19299
/rpt_family="MLT1J1"
complement(19894...20002)
/rpt_family="MIR"
repeat_region
complement(20027...20501)
/rpt_family="MLT2CB"
repeat_region
complement(20652...20823)
/rpt_family="LTR7"
repeat_region
complement(21034...21559)
/rpt_family="MLT1E2"
22407...22453
/rpt_family="(TG)n"

```

```

Query Match      74.4%; Score 18.6; DB 87; Length 105207;
Best Local Similarity 84.0%; Pred. No. 84;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gacattgcgaagtcacagaatta 25
||||| ||| ||||| |||||
Db 87086 GACATTGGGCATTCACAGTATTA 87062

RESULT 42
CNS01RGD 146000 bp DNA PRI 06-JUN-2000
LOCUS Human chromosome 14 DNA sequence *** IN PROGRESS *** BAC R-68M10 of
DEFINITION library RPCI-11 from chromosome 14 of Homo sapiens (Human),
complete sequence.
ACCESSION AL157915
VERSION AL157915.2 GI:7710968
KEYWORDS HTG; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Genospecies.
Direct Submission
Submitted (06-JUN-2000) to the EMBL/GenBank/DBJ databases
ON May 5, 2000 this sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continue. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage, etc... even if efforts are made to eliminate these
contaminating sequences. The following BAC sequence is oriented
from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : C-2171B16
Downstream BAC (overlapping the SP6 end) : C-2566J3

```

## Overall quality chart :

```

Range : bases
0 : 4
1 - 9 : 48
10 - 19 : 587
20 - 29 : 1407
30 - 39 : 6005
40 - 49 : 12726
50 - 59 : 16183
60 - 69 : 9706
70 - 79 : 9481
80 - 89 : 22208
90 - 99 : 67645

```

Percentage of bases with a quality value >= 40 : 94 %.

## FEATURES

## source

```

1. .146000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone_lib="RPCI-11"
/clone="R-68M10"

```

```

BASE COUNT 47532 a 27851 c 27348 g 43268 t 1 others
ORIGIN

```

```

Query Match      74.4%; Score 18.6; DB 91; Length 146000;
Best Local Similarity 84.0%; Pred. No. 84;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1 gacattgcgaagtcacagaatta 25
||||| ||| ||||| |||||
Db 133405 GATATAGCGCAAGTCACAGAATAA 133429

```

## RESULT 43

CNS01DUV 155822 bp DNA PRI 31-MAY-2000  
 LOCUS Human chromosome 14 DNA sequence \*\*\* IN PROGRESS \*\*\* BAC C-2566J3  
 DEFINITION of library Caltech-D from chromosome 14 of Homo sapiens (human), complete sequence.  
 ACCESSION AL133370  
 VERSION AL133370.4 GI:8248722  
 KEYWORDS HTG; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 155822)  
 AUTHORS Direct Submission  
 TITLE Submitted (31-MAY-2000) to the EMBL/GenBank/DBJ databases  
 JOURNAL On Jun 5, 2000 this sequence version replaced gi:8217897.  
 COMMENT IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, phage, etc. . . even if efforts are made to eliminate these contaminating sequences. The following BAC sequence is oriented from the T7 to the SP6 end.  
 Upstream BAC (overlapping the T7 end) : R-68M10  
 Downstream BAC (overlapping the SP6 end) : C-3225F7  
 -----  
 Overall quality chart :  
 Range : bases  
 0 :  
 1 - 9 :  
 10 - 19 :  
 20 - 29 : 4  
 30 - 39 : 24  
 40 - 49 : 1372  
 50 - 59 : 8561  
 60 - 69 : 12248  
 70 - 79 : 10165  
 80 - 89 : 27243  
 90 - 99 : 96205  
 -----  
 Percentage of bases with a quality value >= 40 : 99 %.

FEATURES  
source

Location/Qualifiers  
 1. 155822  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="14"  
 /clone\_lib="Caltech-D"  
 /clone="C-2566J3"  
 53628 a 30002 c 27618 g 44574 t

BASE COUNT  
ORIGIN

Query Match 74.4%; Score 18.6; DB 91; Length 155822;  
 Best Local Similarity 84.0%; Pred. No. 84;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gacattgtgcgaagtccagaatta 25  
 |||||  
 Db 11929 GATATAGGCGCAAGTCACAGAATAA 11953

## RESULT 44

AC016198/c  
 LOCUS Homo sapiens clone RP11-24K5, WORKING DRAFT SEQUENCE, 9 unordered  
 DEFINITION pieces.  
 AC016198  
 VERSION AC016198.3 GI:10045525  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 161474)

## AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

## TITLE

Homo sapiens, clone RP11-24K5

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 161474)

## AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Bouckgalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferrelira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehocsky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye.W.J., Zimmer,A. and Zody,M.

## TITLE

Direct Submission

## JOURNAL

Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Sep 9, 2000 this sequence version replaced gi:6715884.

All repeats were identified using RepeatMasker:

Smit, A.F.A. &amp; Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4488

Center clone name: 24\_K\_5

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 155685 bases at least Q40

Consensus quality: 158433 bases at least Q30

Consensus quality: 159540 bases at least Q20

Insert size: 164000; agarose-fp

Insert size: 160674; sum-of-contigs

Quality coverage: 5.0 in Q20 bases; agarose-fp

Quality coverage: 5.1 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 23310: contig of 23310 bp in length  
 23311 23410: gap of 100 bp  
 23411 27284: contig of 3874 bp in length  
 27285 27384: gap of 100 bp  
 27385 32914: contig of 5530 bp in length  
 32915 33014: gap of 100 bp  
 33015 38790: contig of 5776 bp in length  
 38791 38890: gap of 100 bp  
 38891 46686: contig of 7796 bp in length  
 46687 46786: gap of 100 bp  
 46787 97752: contig of 50966 bp in length  
 97753 97852: gap of 100 bp  
 97853 111746: contig of 13894 bp in length  
 111747 111846: gap of 100 bp  
 111847 151106: contig of 39260 bp in length  
 151107 151206: gap of 100 bp

151207	161474	contig of 10268 bp in length.
FEATURES		
source		
1. 161474		
/organism="Homo sapiens"		
/db_xref="taxon:9606"		
/clone_rpl="24K5"		
/clone_lib="RPC1-11 Human Male BAC"		
1. 23310		
misc_feature		
clone="assembly_fragment"		
clone_end:SP6		
vector_side:left		
23411..27284		
note="assembly_fragment"		
27385..32914		
note="assembly_fragment"		
33015..38790		
note="assembly_fragment"		
38891..46686		
note="assembly_fragment"		
46787..97752		
note="assembly_fragment"		
97853..111746		
note="assembly_fragment"		
111847..151106		
note="assembly_fragment"		
151207..161474		
note="assembly_fragment"		
clone_end:17		
vector_side:right		
50187 a 30235 c 30631 g 49621 t	800 others	
BASE COUNT		
ORIGIN		
Query Match	74.4%	Score 18.6; DB 64; Length 161474;
Best Local Similarity	84.0%	Pred. No. 84;
Matches	21; Conservative	0; Mismatches 4; Indels 0; Gaps 0;
QY	1	gacatttcgcgaagtcacagaatta 25
Db	108923	GATATAGCGCAAGTCACAGATAA 108899
RESULT	45	
LOCUS	AC021065/c	
DEFINITION	Homo sapiens, WORKING DRAFT SEQUENCE, 8 unordered pieces.	
ACCESSION	AC021065	
VERSION	AC021065.6	GI:11079329
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE		
AUTHORS		
1 (bases 1 to 179340)		
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,		
Alisbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,		
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,		
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,		
Burch,P., Burckett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,		
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,		
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,		
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,		
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,		
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,		
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,		
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,		
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,		
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,		
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,		
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,		
Hollins,B., Homsí,F., Howard,S., Huber,J., Huiyá,S., Hume,J.,		
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S.,		
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,		

FEATURES \* 178103 179340: contig of 1238 bp in length.  
Location/Qualifiers  
source 1. .179340  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 51466 a 36447 c 37194 g 53513 t 720 others  
ORIGIN

Query Match 74.4%; Score 18.6; DB 66; Length 179340;  
Best Local Similarity 84.0%; Pred. No. 84;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 gacattgtcgcaagtcacagaatta 25  
||||||| ||| ||||| |||||  
Db 87085 GACATTGTGGCATTTCACAGTATTA 87061

Search completed: October 9, 2001, 12:08:32  
Job time: 3627 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:08:05 ; Search time 470.56 Seconds  
(without alignments)  
33.359 Million cell updates/sec

Title: US-09-396-196f-1  
Perfect score: 25  
Sequence: 1 gacattgcgaagtcacagaatta 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N.Geneseq\_0601:\*

- 1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT.\*
- 2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT.\*
- 3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT.\*
- 4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT.\*
- 5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT.\*
- 6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT.\*
- 7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT.\*
- 8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT.\*
- 9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT.\*
- 10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT.\*
- 11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT.\*
- 12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT.\*
- 13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT.\*
- 14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT.\*
- 15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT.\*
- 16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT.\*
- 17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT.\*
- 18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT.\*
- 19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT.\*
- 20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT.\*
- 21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT.\*
- 22: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	20 AAX01303	E. coli biotin syn
2	25	100.0	1084	10 AAN91329	E.coli Bio B gene.
3	25	100.0	1121	7 AAN60496	Sequence encoding
4	25	100.0	5872	15 AAQ62386	Biotin-biosynthesi
5	17.8	71.2	3083	12 AAQ11851	Glutamate receptor
6	17.6	70.4	512	20 AAV88744	Est clone HK26. H
7	17.6	70.4	1041	10 AAN91327	E.coli Bio B gene
8	17.2	68.8	322	16 AAT91329	Human gene signatu
9	17.2	68.8	902	21 AA256756	Human transmembran
10	17	68.0	17	14 AAQ56307	BioB DNA primer.
11	16.6	66.4	1374	15 AAQ68434	Maize Hml cDNA. Z

C 12	16.6	66.4	1374	16	AAQ99463	Maize Hml gene cDN
C 13	16.6	66.4	2157	8	AAN70438	Streptococcus pyog
C 14	16.6	66.4	3717	21	AAC48934	Arabidopsis thalia
C 15	16.6	66.4	5198	15	AAQ84833	Maize Hml gene. Z
C 16	16.6	66.4	5198	16	AAQ99462	Maize Hml genomic
C 17	16.2	64.8	386	21	AAC24233	Human secreted pro
C 18	16.2	64.8	1150	21	AAC39577	Arabidopsis thalia
C 19	16.2	64.8	1771	20	AAZ30602	Mouse integrin alp
C 20	16.2	64.8	2114	14	AAQ51236	Plant NADPH cytoch
C 21	16.2	64.8	2697	20	AAX78082	A. gossypii GUA2 D
C 22	16.2	64.8	2879	18	AAT67318	Xenopus beta-signa
C 23	16.2	64.8	2989	16	AAT02799	Human glutamate re
C 24	16.2	64.8	2989	16	AAT02800	Human glutamate re
C 25	16.2	64.8	3070	14	AAQ62694	Human GluR3B recep
C 26	16.2	64.8	3251	21	AAT21307	Human low adenosin
C 27	16.2	64.8	3251	21	AAAS1585	Human adenosine re
C 28	16.2	64.8	40298	21	AAF21311	Human low adenosin
C 29	16.2	64.8	40298	21	AAAS1589	Human adenosine re
C 30	16	64.0	420	18	AAT67395	H. pylori surface
C 31	16	64.0	423	19	AAX30422	H. pylori cellular
C 32	16	64.0	477	18	AAT68284	H. pylori transmem
C 33	16	64.0	537	21	AAC36423	Arabidopsis thalia
C 34	16	64.0	1157	21	AAA81657	N. meningitidis pa
C 35	16	64.0	2973	21	AAA29208	Human topoisomeras
C 36	16	64.0	4812	21	AAA13178	Vitellogenin encod
C 37	16	64.0	5994	21	AAA94634	Human TUB 5' regio
C 38	16	64.0	5995	18	AAT96641	Human TUB locus 5'
C 39	16	64.0	6688	19	AAV58288	S. pyogenes SP-26-
C 40	16	64.0	1230025	20	AAX91990	Nucleotide sequenc
C 41	15.8	63.2	60	14	AAQ51028	Fragment of the hu
C 42	15.8	63.2	253	15	AAQ76727	Human genome fragm
C 43	15.8	63.2	628	21	AAC41591	Arabidopsis thalia
C 44	15.8	63.2	629	21	AAC39882	Arabidopsis thalia
C 45	15.8	63.2	837	21	AAF08144	Fusarium venenatum

## ALIGNMENTS

RESULT 1

AAX01303

ID AAX01303 standard; DNA; 1041 BP.

XX

AC AAX01303;

XX

DT 12-APR-1999 (first entry)

XX

XX E. coli biotin synthetase (BioB) coding sequence.

DE

DE DAP aminotransferase; diaminopelargonic acid; transgenic plant;

KW biotin synthase; biotin production; vitamin H; BioB; ss.

KW

XX Escherichia coli.

OS

XX

PN US5869719-A.

XX

PD 09-FEB-1999.

XX

PF 30-APR-1997; 97US-0846338.

XX

PR 30-APR-1997; 97US-0846338.

PR

XX 08-MAR-1995; 95US-0401068.

XX

XX (NOVS ) NOVARTIS FINANCE CORP.

PA Patton DA;

XX

XX WPI; 1999-152902/13.

DR

DR P-PSDB; AAW73906.

XX

XX Transgenic plants with high biotin levels - transformed with DNA

PT encoding di:amino-pelargonic acid amino-transferase or biotin

PT synthase

XX Example 2; Column 37-40; 34pp; English.

PS This sequence encodes the E. coli biotin synthetase (BioB). The gene can

CC be used in the transgenic plant of the invention. The transgenic plant,

CC plant cell or plant tissue is transformed with a chimeric gene encoding

CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and

CC produces more biotin than a non-transgenic plant, cell or tissue. The

CC plant is used as an improved dietary source of biotin (vitamin H) for

CC humans or animals.

XX Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

SQ

Query Match 100.0%; Score 25; DB 20; Length 1041;

Best Local Similarity 100.0%; Pred. No. 0.0097;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaatta 25

Db 21 gacattgtcgcaagtcacagaatta 45

|||||

RESULT 2

AA091329

ID AAN91329 standard; DNA; 1084 BP.

XX

AC AAN91329;

XX

DT 15-FEB-1990 (first entry)

XX

DE E. coli Bio B gene.

XX

XX E. coli; Bio B gene; biotin.

KW Escherichia coli.

XX

OS

XX

PH Key Location/Qualifiers

FT CDS 24..1084

FT /\*tag=a

XX

PN GB2216530-A.

PD

XX

PD 11-OCT-1989.

XX

PF 17-MAR-1989; 89GB-0006210.

XX

PR 22-MAR-1988; 88GB-0006804.

XX

PR 17-MAR-1989; 89GB-0006210.

XX

PA (UKAG-) UK MIN. AGRIC. FISH.

XX

PI Pearson BM, McKee RA;

XX

DR WPI; 1989-295085/41. P-PSDB P91392

XX

XX Plasmid contg. gene(s) for expression of biotin synthetase enzymes

PT - derived from E. coli and capable of replication and expression in other

PT microorganisms, esp. yeast.

XX

PS Table 3; page 33-4; 52pp; English.

XX

CC The gene can be used in a plasmid for expression of enzymes of the biotin

CC synthetic pathway. Pref. control sequences for expression in S. cerevisiae

CC are plasmids pNA91, pNA36c, pKV49 and pCK495, and plasmid pCK965 for

CC Lactobacillus. Insertion of bio B improves biotin yields in

CC microorganisms which export biotin, or enables growth in media contg.

CC little or no biotin of organisms unable to synthesise biotin for their

CC own use.

XX

SQ Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;

Best Local Similarity 100.0%; Pred. No. 0.0098;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaatta 25

Db 44 gacattgtcgcaagtcacagaatta 68

|||||

RESULT 3

AA060496

ID AAN60496 standard; DNA; 1121 BP.

XX

AC AAN60496;

XX

DT 17-OCT-1991 (first entry)

XX

XX Sequence encoding biotin synthesising enzyme.

DE Biotin synthetic enzyme; E. coli; desthiobiotin; ds.

XX

KW

XX

PH Key Location/Qualifiers

FT CDS 42..1082

FT /\*tag= a

XX

PN JP61149091-A.

PD

XX

PD 07-JUL-1986.

XX

PF 24-DEC-1984; 84JP-0272605.

XX

PR 24-DEC-1984; 84JP-0272605.

XX

XX (NIPS ) NIPPON SODA KK.

XX

XX WPI; 1986-216622/33.

DR P-PSDB; AAP60536.

XX

PT Double stranded DNA encoding biotin synthesising enzyme -

PT comprises transformed mutant E. coli strain contg. cyclic doubled

PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.

XX

PS Disclosure; Page 534; 23pp; Japanese.

XX

CC The sequence may be expressed by a transformed E. coli host, cultured

CC in a medium containing desthiobiotin.

XX

SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;

Best Local Similarity 100.0%; Pred. No. 0.0098;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaatta 25

Db 62 gacattgtcgcaagtcacagaatta 86

|||||

RESULT 4

AA062386

ID AA062386 standard; DNA; 5872 BP.

XX

AC AA062386;

XX

DT 16-NOV-1994 (first entry)

XX

XX Biotin-biosynthesis genes contg. plasmid pB030A-15/9.

DE

XX Biotin; expression; enterobacteria; vitamin H; synthesis;

KW plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;

KW promoter ptac; biotin synthase; KAPA synthase;

KW 8-amino-7-oxononanoate synthase; pimeloyl-CoA; DFB synthase;

KW dethiobiotin synthase; DAPA synthase;  
 KW S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;  
 KW seborrhea; dermatitis; ds.  
 XX Escherichia coli DSM498.  
 OS  
 XX  
 FT FT Location/Qualifiers  
 FT FH promoter  
 FT FT 1..96  
 FT FT /tag= a  
 FT FT /function= "promoter ptac"  
 FT FT /evidence= EXPERIMENTAL  
 FT FT 23..28  
 FT FT /tag= b  
 FT FT /standard\_name= "promoter ptac"  
 FT FT 45..50  
 FT FT /tag= c  
 FT FT /evidence= EXPERIMENTAL  
 FT FT /standard\_name= "promoter ptac"  
 FT FT 105..109  
 FT FT /tag= d  
 FT FT /evidence= EXPERIMENTAL  
 FT FT /standard\_name= "bioB RBS no. 9"  
 FT FT 117..1157  
 FT FT /tag= e  
 FT FT /product= "biotin synthase"  
 FT FT /evidence= EXPERIMENTAL  
 FT FT /gene= "bioB"  
 FT FT /number= 1  
 FT FT 1141..1146  
 FT FT /tag= f  
 FT FT /standard\_name= "bioF RBS"  
 FT FT 1154..2311  
 FT FT /tag= g  
 FT FT /EC\_number= 2.3.1.47  
 FT FT /product= "KAPA synthase"  
 FT FT /evidence= EXPERIMENTAL  
 FT FT /gene= "bioF"  
 FT FT /number= 2  
 FT FT /standard\_name= "8-amino-7-oxononanoate synthase"  
 FT FT 2284..2288  
 FT FT /tag= h  
 FT FT /standard\_name= "bioC RBS"  
 FT FT 2295..3050  
 FT FT /tag= i  
 FT FT /function= "involved in pimeloyl-CoA synthesis"  
 FT FT /product= "protein"  
 FT FT /gene= "bioC"  
 FT FT /number= 3  
 FT FT 3030..3033  
 FT FT /tag= j  
 FT FT /standard\_name= "bioD RBS"  
 FT FT 3043..3753  
 FT FT /tag= k  
 FT FT /EC\_number= 6.3.3.3  
 FT FT /product= "DTB synthase"  
 FT FT /evidence= EXPERIMENTAL  
 FT FT /gene= "bioD15"  
 FT FT /number= 4  
 FT FT /standard\_name= "dethiobiotin synthase"  
 FT FT 3712..3750  
 FT FT /tag= l  
 FT FT /note= "bioD15 substitution"  
 FT FT 3742..3746  
 FT FT /tag= m  
 FT FT /standard\_name= "bioA RBS"  
 FT FT 3750..5039  
 FT FT /tag= n  
 FT FT /EC\_number= 2.6.1.62  
 FT FT /product= "DAPA synthase"  
 FT FT /evidence= EXPERIMENTAL  
 FT FT /gene= "bioA"  
 FT FT /number= 5  
 FT FT /standard\_name= "S-adenosyl-L-methionine: 8-amino-

FT RBS 5088..5093  
 FT FT /tag= o  
 FT FT /standard\_name= "ORFI RBS"  
 FT FT 5098..5574  
 FT FT /tag= p  
 FT FT /function= "unknown, involved in biotin synthesis"  
 FT FT /product= "protein"  
 FT FT /evidence= EXPERIMENTAL  
 FT FT /gene= "ORFI"  
 FT FT /number= 6  
 FT FT 5583..5644  
 FT FT /tag= q  
 FT FT /standard\_name= "rho-independent transcriptional  
 FT FT terminator"  
 FT FT 5583..5605  
 FT FT /tag= r  
 FT FT  
 PN WO9408023-A.  
 PD 14-APR-1994.  
 XX  
 XX 01-OCT-1993; 93WO-EP02688.  
 XX  
 XX 02-OCT-1992; 92CH-0003124.  
 PR 15-JUL-1993; 93CH-0002134.  
 XX  
 XX (LONZ ) LONZA AG.  
 XX  
 XX Birch O, Brass J, Fuhrmann M, Shaw N;  
 XX WPI; 1994-135587/16.  
 DR P-PSDB; AARS1883, AARS1884, AARS1885, AARS1886, AARS1887, AARS3121.  
 DR  
 XX Biotechnological biotin prodn. using enterobacterial biotin-gene  
 PT - providing vitamin H in high yield  
 PT  
 XX Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.  
 XX  
 CC The sequence is derived from plasmid pB030A-15/9 contg. the  
 CC bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis  
 CC of biotin, arranged in a transcription unit. Microorganisms  
 CC contg. these DNA fragments or plasmids may be used in the prodn.  
 CC of biotin. Biotin (Vitamin H) may prevent seborrhea, dermatitis,  
 CC loss of appetite and tiredness.  
 XX  
 SQ Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;

Query Match 100.0%; Score 25; DB 15; Length 5872;  
 Best Local Similarity 100.0%; Pred. No. 0.012;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaatta 25  
 |||||  
 Db 137 gacattgtcgcaagtcacagaatta 161

RESULT 5  
 AAQ11851/C  
 ID AAQ11851 standard; DNA; 3083 BP.  
 XX  
 XX AC AAQ11851;  
 XX  
 XX 31-JUL-1991 (first entry)  
 XX  
 XX Glutamate receptor 3.  
 KW Glutamate receptor 3; probe; ligand; drug screening; ss.  
 XX Rattus rattus.  
 XX  
 XX Key Location/Qualifiers

```

FT CDS 167..2830
FT /*tag= a
FT /product= GR3
FT sig_peptide 167..232
FT /*tag= b
FT mat_peptide 233..2830
FT /*tag= c
XX
XX WO9106648-A.
XX
XX 16-MAY-1991.
XX
XX 25-OCT-1990; 90WO-US06153.
XX
XX 27-OCT-1989; 89US-0428116.
XX
XX (SALK ) SALK INST FOR BIOL STUD.
XX
XX Heinemann SF, Boulter JR, Hollmann M, Bettler B, Jensen JE;
XX
XX WPI; 1991-164197/22.
XX
XX P-PSDB; AAR11991.
XX
XX Glutamate receptors - used to screen for functional ligands and
XX identify and isolate further receptors
XX
XX Disclosure; Fig 4; 109pp; English.
XX
XX GluR3 cDNA was isolated from a rat forebrain cDNA using a
XX low-stringency screening protocol and a radiolabelled fragment of the
XX GluR1 cDNA as probe. The cDNA is deposited (ATCC 68133).
XX The gene and protein can be used in drug screening, to
XX determine whether a substance is a functional ligand for the
XX receptor by monitoring ion channel activity.
XX See also AAQ11845-855.
XX
XX Sequence 3083 BP; 877 A; 679 C; 735 G; 792 T; 0 other;
XX
XX
XX Query Match 71.2%; Score 17.8; DB 12; Length 3083;
XX Best Local Similarity 90.5%; Pred. No. 28;
XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX
XX 3 cattgtgcgaagtccacagaatt 23
XX |||||||||
XX Db 3034 CATTGTCGCAAGTCTCAGAGT 3014
XX
XX
XX RESULT 6
XX AAV88744
XX ID AAV88744 standard; cDNA; 512 BP.
XX
XX AAV88744;
XX
XX 12-FEB-1999 (first entry)
XX
XX EST clone HK26.
XX
XX Expressed sequence tag; secreted protein; haematopoiesis regulator;
XX tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
XX chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
XX receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
XX Homo sapiens.
XX
XX WO9845437-A2.
XX
XX 15-OCT-1998.
XX
XX 10-APR-1998; 98WO-US06956.
XX
XX 10-APR-1997; 97US-0837312.
XX
XX

```

```

PA (GEM ) GENETICS INST INC.
XX
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
XX Racie LA, Spaulding V, Treacy M;
XX
XX WPI; 1999-070078/06.
XX
XX New polynucleotides encoding human secreted proteins - derived from
XX e.g. human blood, kidney, foetal lung, placenta, testes, brain,
XX ovary, pituitary, retina and colon cDNA libraries
XX
XX Claim 1; Page 498; 64pp; English.
XX
XX The present sequence represents an expressed sequence tag (EST), and is
XX a polynucleotide of the invention. The polynucleotides of the invention
XX are all secreted EST sequences isolated from a variety of human tissue
XX sources. The EST sequences and proteins encoded by them are predicted to
XX have useful biological activities which would make them suitable for
XX treating, preventing or ameliorating medical conditions in humans and
XX animals, although no supporting data is given. Suggested activities
XX include nutritional activity, immune stimulating or suppressing activity,
XX haematopoiesis regulating activity, tissue growth activity,
XX activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX activity, cadherin/tumour invasion suppressor activity, tumour inhibition
XX activity. The EST sequences are also stated to be useful for gene
XX therapy.
XX
XX Sequence 512 BP; 160 A; 104 C; 96 G; 152 T; 0 other;
XX
XX
XX Query Match 70.4%; Score 17.6; DB 20; Length 512;
XX Best Local Similarity 83.3%; Pred. No. 28;
XX Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
XX
XX 1 gacattgtgcgaagtccacagaatt 24
XX | ||||| | |||||||||
XX Db 79 gtcattgtgccagtcacagaatt 102
XX
XX
XX RESULT 7
XX AAN91327
XX ID AAN91327 standard; DNA; 1041 BP.
XX
XX AAN91327;
XX
XX 15-FEB-1990 (first entry)
XX
XX E.coli Bio B gene modified for yeast.
XX
XX E.coli; modified Bio B gene; biotin.
XX
XX Escherichia coli.
XX
XX GB2216530-A.
XX
XX 11-OCT-1989.
XX
XX 17-MAR-1989; 89GB-0006210.
XX
XX 22-MAR-1988; 88GB-0006804.
XX
XX 17-MAR-1989; 89GB-0006210.
XX
XX (UKAG-) UK MIN. AGRIC. FISH.
XX
XX Pearson BM, McKee RA;
XX
XX WPI; 1989-295085/41.
XX
XX Plasmid contg. gene(s) for expression of biotin synthetase enzymes
XX - derived from E.coli and capable of replication and expression in other
XX microorganisms, esp. yeast.
XX

```



XX  
KW Hm1 gene; HC toxin; race-specific resistance; disease-resistance;  
KW *Cochliobolus carbonum*; selectable marker; transformation;

CC A DNA sequence is claimed having 90% translational homology to  
 CC genomic or cDNA sequence of maize Hm1 gene (AAQ99462 & AAQ99463). The  
 CC gene Hm1 controls resistance to Cochliobolus carbonum Nelson race 1.  
 CC The Hm1 gene can be used in conjunction with HC-toxin in a  
 CC selectable marker system for use in maize transformation.  
 CC This gene can also be inserted into the genomes of maize varieties  
 CC which lack the gene to produce transformants that are resistance  
 CC to disease caused by C. carbonum. The Hm1 resistance gene is one of  
 CC a family of homologous disease resistance genes in maize and other  
 CC crops. The invention encompasses methods for identification of  
 CC homologous disease resistance genes by using this gene or a  
 CC fragment of it as a probe.

XX SQ Sequence 1374 BP; 287 A; 405 C; 432 G; 250 T; 0 other;

Query Match 66.4%; Score 16.6; DB 16; Length 1374;

Best Local Similarity 82.6%; Pred. No. 94; Indels 0; Gaps 0;  
 Matches 19; Conservative 0; Mismatches 4;

QY 2 acattgtcgcaagtcacagaatt 24  
 ||||| ||||| ||||| |||||  
 Db 1329 AANTATCGCAAGTCATATAATT 1307

RESULT 13  
 AAN70438/c  
 ID AAN70438 standard; DNA; 2157 BP.

XX AC AAN70438;

XX DT 26-FEB-1991 (first entry)

XX DE Streptococcus pyogenes antitumor SAGP protein.

XX KW antitumor; SAGP gene; acid glycoprotein; ss.

XX OS Streptococcus pyogenes.

XX FH Key Location/Qualifiers  
 XX CDS 40..1219  
 XX FT /\*tag= a  
 XX FT /label=SAGP gene

XX PN EP230777-A.

XX PD 05-AUG-1987.

XX PF 23-DEC-1986; 86EP-0310075.

XX PR 21-AUG-1986; 86JP-0194077.

XX PR 28-DEC-1985; 85JP-0298014.

XX PA (OHGE-) OHGEN RES LAB LTD.

XX PI Kanaoka M, Negoro T, Kawanaka C, Agui H;

XX DR WPI; 1987-215081/31.

XX DR P-PSDB; AAP70278.

XX PT New DNA sequence coding for streptococcal antitumor protein - and  
 XX PT derived expression vectors and transformed E. coli strains.

XX PS Disclosure; fig 2; 16pp; English.

XX CC This sequence is encoded by plasmid pSPI and may be expressed in E.  
 XX CC coli JMI03 to give large pure quantities of the antitumor protein.

XX SQ Sequence 2157 BP; 682 A; 416 C; 434 G; 625 T; 0 other;

Query Match 66.4%; Score 16.6; DB 8; Length 2157;

Best Local Similarity 82.6%; Pred. No. 99;  
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 cattgtcgcaagtcacagaatta 25  
 ||||| ||||| ||||| |||||  
 Db 958 CATTCGCTAAGTGACAGAGTAA 936

RESULT 14

AAC48934

ID AAC48934 standard; DNA; 3717 BP.

XX AC AAC48934;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 59311.

XX KW Hybridisation assay; genetic mapping; gene expression control;

XX KW protein identification; signal transduction pathway;

XX KW metabolic pathway; promoter; termination sequence; ss.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 06-APR-1999; 99US-0128234.

XX PR 08-APR-1999; 99US-0128714.

XX PR 16-APR-1999; 99US-0129845.

XX PR 19-APR-1999; 99US-0130077.

XX PR 21-APR-1999; 99US-0130449.

XX PR 23-APR-1999; 99US-0130510.

XX PR 28-APR-1999; 99US-0130891.

XX PR 30-APR-1999; 99US-0131449.

XX PR 30-APR-1999; 99US-0132048.

XX PR 04-MAY-1999; 99US-0132484.

XX PR 05-MAY-1999; 99US-0132485.

XX PR 06-MAY-1999; 99US-0132486.

XX PR 06-MAY-1999; 99US-0132487.

XX PR 07-MAY-1999; 99US-0132863.

XX PR 11-MAY-1999; 99US-0134256.

XX PR 14-MAY-1999; 99US-0134218.

XX PR 14-MAY-1999; 99US-0134221.

XX PR 14-MAY-1999; 99US-0134370.

XX PR 18-MAY-1999; 99US-0134768.

XX PR 19-MAY-1999; 99US-0134941.

XX PR 20-MAY-1999; 99US-0135124.

XX PR 21-MAY-1999; 99US-0135353.

XX PR 24-MAY-1999; 99US-0135629.

XX PR 25-MAY-1999; 99US-0136021.

XX PR 27-MAY-1999; 99US-0136392.

XX PR 28-MAY-1999; 99US-0136782.

XX PR 01-JUN-1999; 99US-0137222.

XX PR 03-JUN-1999; 99US-0137528.

XX PR 04-JUN-1999; 99US-0137502.

XX PR 07-JUN-1999; 99US-0137724.

XX PR 08-JUN-1999; 99US-0138094.

XX PR 10-JUN-1999; 99US-0138540.

XX PR 10-JUN-1999; 99US-0138847.

XX PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140354.  
PR 28-JUN-1999; 99US-0140695.  
PR 29-JUN-1999; 99US-0140823.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147935.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 22-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 25-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 25-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 66.4%; Score 16.6; DB 21; Length 3717;  
Best Local Similarity 82.6%; Pred. No. 1, 1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gacattgtcgaagtacagaat 23  
||||| ||||| |||||  
Db 2935 gacattgttcgaagtacagaat 2957



RESULT 15  
 AAQ68433/c  
 ID AAQ68433 standard; DNA; 5198 BP.  
 XX  
 AC AAQ68433;  
 XX  
 DT 26-JAN-1995 (first entry)  
 XX  
 DE Maize Hml gene.  
 XX  
 KW Hml gene; HC toxin; race-specific resistance; disease-resistance;  
 KW Cochliobolus carbonum; selectable marker; transformation;  
 KW transposon tagging; ds.  
 XX  
 OS Zea mays.  
 XX  
 PN W09413825-A.  
 XX  
 PD 23-JUN-1994.  
 XX  
 PF 14-DEC-1993; 93WO-US12146.  
 XX  
 PR 15-DEC-1992; 92US-0995658.  
 XX  
 PA (PION-) PIONEER HI-BRED INT INC.  
 XX  
 PI Briggs SP, Johal GS;  
 XX  
 DR WPI; 1994-217898/26.  
 XX  
 PT Hml gene conferring race-resistance to Cochliobolus carbonum to  
 PT maize - for use as a selectable marker for transformed maize  
 PT cells  
 XX  
 PS Disclosure; Page 14-16; 19pp; English.  
 XX  
 CC Transposon mutagenesis was used to tag, clone and characterize the  
 CC maize Hml gene. Genomic and cDNA sequences of the Hml gene are  
 CC provided in AAQ68433 and AAQ68434, respectively.  
 CC  
 SQ Sequence 5198 BP; 1376 A; 1160 C; 1135 G; 1527 T; 0 other;  
  
 Query Match 66.4%; Score 16.6; DB 15; Length 5198;  
 Best Local Similarity 82.6%; Pred. No. 1.1e+02;  
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
 QY 2 acattgtcgcaagtcacagaatt 24  
 | ||| ||||| ||||| |||||  
 Db 4891 AAATTATCGCAAGTCATAAAATT 4869  
  
 RESULT 16  
 AAQ99462/c  
 ID AAQ99462 standard; DNA; 5198 BP.  
 XX  
 AC AAQ99462;  
 XX  
 DT 30-DEC-1995 (first entry)  
 XX  
 DE Maize Hml genomic DNA.  
 XX  
 KW Hml gene; Cochliobolus carbonum Nelson race 1; fungus;  
 KW fungal disease toxin; resistance; ss.  
 XX  
 OS Zea mays.  
 XX  
 PN W09507989-A.  
 XX  
 PD 23-MAR-1995.  
 XX  
 PF 19-SEP-1994; 94WO-US10497.

XX 17-SEP-1993; 93US-0123761.  
 PR (PION-) PIONEER HI-BRED INT INC.  
 XX  
 PA Briggs SP, Johal GS;  
 XX  
 PI WPI; 1995-131357/17.  
 XX  
 DR Disease resistance gene Hml from maize - is used to confer  
 XX resistance to Cochliobolus carbonum.  
 PT  
 PT Claim 2; Page 20-22; 26pp; English.  
 XX  
 CC A DNA sequence is claimed having 90% translational homology to  
 CC genomic or cDNA sequence of maize Hml gene (AAQ99462 & AAQ99463). The  
 CC gene Hml controls resistance to Cochliobolus carbonum Nelson race 1.  
 CC The Hml gene can be used in conjunction with HC-toxin in a  
 CC selectable marker system for use in maize transformation.  
 CC This gene can also be inserted into the genomes of maize varieties  
 CC which lack the gene to produce transformants that are resistance  
 CC to disease caused by C. carbonum. The Hml resistance gene is one of  
 CC a family of homologous disease resistance genes in maize and other  
 CC crops. The invention encompasses methods for identification of  
 CC homologous disease resistance genes by using this gene or a  
 CC fragment of it as a probe.  
 XX  
 SQ Sequence 5198 BP; 1375 A; 1159 C; 1136 G; 1528 T; 0 other;  
  
 Query Match 66.4%; Score 16.6; DB 16; Length 5198;  
 Best Local Similarity 82.6%; Pred. No. 1.1e+02;  
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
 QY 2 acattgtcgcaagtcacagaatt 24  
 | ||| ||||| ||||| |||||  
 Db 4891 AAATTATCGCAAGTCATAAAATT 4869  
  
 RESULT 17  
 AAC24233/c  
 ID AAC24233 standard; cDNA; 386 BP.  
 XX  
 AC AAC24233;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein 5' EST, SEQ ID NO: 28308.  
 XX  
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000EP-0200610.  
 XX  
 PR 26-FEB-1999; 99US-0122487.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI; 2000-500381/45.  
 XX  
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 PS Claim 1; SEQ ID 28308; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA<sup>+</sup> RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.

XX Sequence 386 BP; 121 A; 55 C; 58 G; 144 T; 8 other;

Query Match 64.8%; Score 16.2; DB 21; Length 386;  
 Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ttgtcgaagtccacagaatta 25  
 ||||| ||||| ||||| |||||  
 Db 240 TTGTTGAAACTCACAGAATTA 220

RESULT 18  
 AAC39577/c  
 ID AAC39577 standard; DNA; 1150 BP.  
 XX  
 AC AAC39577;  
 XX  
 DT 17-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 25145.  
 XX  
 KW Hybridisation assay; genetic mapping; gene expression control;  
 KW protein identification; signal transduction pathway;  
 KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

```

PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.

PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 64.8%; Score 16.2; DB 21; Length 1150;
Best Local Similarity 85.7%; Pred No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gacattgtcgcagtcacaga 21
    ||||| ||||| ||| |||
Db 79 GACATAGTCGCAACTCAAGA 59

RESULT 19
AAZ30602
ID AAZ30602 standard; DNA; 1771 BP.
XX
AC AAZ30602;
XX
DT 18-JAN-2000 (first entry)
XX
DE Mouse integrin alpha 4 coding sequence.
XX
KW Human; integrin; antisense; oligonucleotide; inhibition; expression;
KW very late antigen; CD49d; CD29; cell surface; leucocyte; adhesion;
KW vascular endothelial cell; vascular endothelium; migration; inflammation;
KW atherosclerosis; allergy; asthma; rheumatoid arthritis; tumor;
KW metastasis; circulatory system; autoimmune disease; Grave's disease;
KW Hashimoto's thyroiditis; encephalomyelitis; multiple sclerosis; ds.
XX
OS Mus sp.
XX
PN US5968826-A.
XX
PD 19-OCT-1999.
XX
PF 05-OCT-1998; 98US-0166203.
XX
PR 05-OCT-1998; 98US-0166203.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Cowsett LM, Condon TP;
XX
DR WPI; 1999-590416/50.
DR P-PSDB; AAY44154.
XX
PT Antisense inhibition of integrin alpha4 expression useful for treating
PT inflammatory diseases such as atherosclerosis, allergies, asthma and
PT arthritis -
XX
PS Example 12; Column 71-74; 40pp; English.
XX
CC This sequence represents the coding region of the mouse integrin
CC alpha4 gene. The invention relates to the generation of antisense
CC oligonucleotides targeted to the integrin alpha4 gene which are used
CC for inhibiting expression of the integrin alpha4 mRNA or protein.
CC Integrin alpha4 is a component of Very Late Antigen (VLA)-4 (also
CC called alpha4beta1 and CD49d/CD29). VLA-4 is expressed on the cell
CC surfaces of leucocytes and vascular endothelial cells and mediates the
CC adhesion of leucocytes to the vascular endothelium prior to migration
CC into the surrounding tissues. This migration is an essential step in
CC inflammation and hence VLA-4 (and consequently integrin alpha4) is a
CC potential therapeutic target for treating inflammatory disorders and
CC the damaging effects of excessive inflammation. These disorders include

```

CC atherosclerosis, allergies, asthma, rheumatoid arthritis and tumor cell  
 CC metastasis (VLA-4 is involved in migration of the tumor cells through  
 CC the extracellular matrix into the circulatory system). VLA-4 is also  
 CC involved in a number of autoimmune diseases such as Grave's disease,  
 CC Hashimoto's thyroiditis, encephalomyelitis (EAE), multiple sclerosis.  
 CC VLA-4 may also be involved in promoting adhesion (i.e. retention) of  
 CC hemopoietic stem cells in bone-marrow and in allograft rejection.  
 XX

SQ Sequence 1771 BP; 371 A; 471 C; 519 G; 410 T; 0 other;

Query Match 64.8%; Score 16.2; DB 20; Length 1771;

Best Local Similarity 85.7%; Pred. NO. 1.5e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 attgtcgcaagtccacagaatt 24

|||||

Db 306 agtgaggcaagtccacagaatt 326

RESULT 20

AAQ51236

ID AAQ51236 standard; cDNA; 2114 BP.

XX AAQ51236;

XX

DT 11-MAY-1994 (first entry)

XX

DE Plant NADPH cytochrome P450 reductase (ara B).

XX NADPH cytochrome P450 reductase; functional complementation;

KW identification; ss.

KW

XX Arabidopsis thaliana.

OS

XX

FH Key Location/Qualifiers

FT CDS 36..2114

FT /tag= a

FT /product= NADPH cytochrome P450 reductase.

FT misc\_difference 510..512

FT /tag= c

FT /transl\_except= AGA encodes Lys.

FT misc\_difference 543..545

FT /tag= d

FT /transl\_except= CGG encodes Lys.

FT misc\_difference 663..665

FT /tag= e

FT /transl\_except= CGT encodes Lys.

FT misc\_difference 819..821

FT /tag= f

FT /transl\_except= CGG encodes Lys.

FT misc\_difference 915..917

FT /tag= g

FT /transl\_except= AGA encodes Lys.

FT misc\_difference 963..965

FT /tag= h

FT /transl\_except= CGG encodes Lys.

FT misc\_difference 996..998

FT /tag= i

FT /transl\_except= AGG encodes Lys.

FT misc\_difference 1203..1205

FT /tag= j

FT /transl\_except= AGA encodes Lys.

FT misc\_difference 1230..1232

FT /tag= k

FT /transl\_except= CGA encodes Lys.

FT misc\_difference 1431..1433

FT /tag= l

FT /transl\_except= CGT encodes Lys.

FT misc\_difference 1443..1445

FT /tag= m

FT /transl\_except= CGT encodes Lys.

FT misc\_difference 1484..1486

FT /tag= n  
 FT /transl\_except= AGA encodes Lys.  
 FT misc\_difference 1533..1535  
 FT /tag= o  
 FT /transl\_except= AGA encodes Lys.  
 FT misc\_difference 1623..1625  
 FT /tag= p  
 FT /transl\_except= CGA encodes Lys.  
 FT misc\_difference 1704..1706  
 FT /tag= q  
 FT /transl\_except= AGA encodes Lys.  
 FT misc\_difference 1782..1784  
 FT /tag= p  
 FT /transl\_except= AGA encodes Lys.  
 FT misc\_difference 1788..1790  
 FT /tag= q  
 FT /transl\_except= CGA encodes Lys.  
 FT misc\_difference 1872..1874  
 FT /tag= r  
 FT /transl\_except= CGT encodes Lys.  
 FT misc\_difference 1992..1994  
 FT /tag= s  
 FT /transl\_except= AGG encodes Lys.  
 FT misc\_difference 2004..2006  
 FT /tag= t  
 FT /transl\_except= CGA encodes Lys.  
 FT misc\_difference 2091..2093  
 FT /tag= u  
 FT /transl\_except= AGA encodes Lys.  
 FT misc\_difference 2100..2102  
 FT /tag= v  
 FT /transl\_except= AGA encodes Lys.

WO9321326-A.

28-OCT-1993.

13-APR-1993; 93WO-FR00367.

13-APR-1992; 92FR-0004491.

(ORSA-) ORSAN.

Kazmaier M, Lacroute F, Mignotte-vieux C, Minet M;

Pompon D;

WPI: 1993-351736/44.

P-PSDB; AAR43581.

New DNA encoding plant NADPH cytochrome P450 reductase - cloned

by functional complementation in yeast, also recombinant enzyme

useful in P450 mediated bioconversion processes

Claim 11; Figure 9; 79pp; French.

A new method for determining whether a DNA sequence encodes an NADPH  
 cytochrome P450 reductase involves transforming yeasts with plasmids  
 of a total cDNA bank of plant(s). The yeasts used in the procedure  
 are incapable of producing their own NADPH cytochrome P450  
 reductase. They are then exposed to a cytochrome P450 inhibitor at a  
 level which is lethal to the yeast cells but not to cells which,  
 because of the transformation, now contain an active NADPH  
 cytochrome P450 reductase. Surviving clones are then isolated and  
 plasmid DNA extracted. The gene is inserted into the plasmid at a  
 site which places it under the control of an inducible promoter.

Sequence 2114 BP; 603 A; 400 C; 542 G; 569 T; 0 other;

Query Match 64.8%; Score 16.2; DB 14; Length 2114;

Best Local Similarity 85.7%; Pred. NO. 1.5e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 attgtgcgaagtcacagaatt 24  
 ||||| ||||| ||||| |||||  
 Db 1338 attgtgcgaagtcacagaagt 1358

RESULT 21  
 AAX78082/c  
 ID AAX78082 standard; DNA; 2697 BP.

XX AC AAX78082;  
 XX XX  
 DT 19-AUG-1999 (first entry)  
 XX DE  
 DE A. gossypii GUA2 DNA.  
 XX XX  
 KW Purine biosynthesis; phosphoribosylpyrophosphate synthetase; KPR1; KPR2;  
 KW glutamine-phosphoribosylpyrophosphate amidotransferase; ADE4; GUA1;  
 KW IMP-dehydrogenase; GUA1; GMP-synthetase; GUA2; PRPP synthetase; mouth;  
 KW riboflavin; vitamin B2; treatment; inflammation; throat; mucosal layer;  
 KW skin disorder; ss.  
 XX OS

XX Ashbya gossypii.  
 XX XX  
 FH Key Location/Qualifiers  
 FT CDS 456..2033  
 FT /\*tag= a  
 FT /product= "Guanosine-monophosphate synthetase"

XX EP927761-A2.  
 XX XX  
 PD 07-JUL-1999.  
 XX XX  
 PF 08-DEC-1998; 98EP-0123331.  
 XX XX  
 PR 23-DEC-1997; 97DE-1057755.

XX (BADI ) BASF AG.  
 XX XX  
 PI Hoeffken HW, Jimenez A, Pompejus M, Revuelta Doval JL;  
 PI Santos Garcia MA, Seulberger H;  
 XX XX  
 DR WPI; 1999-373465/32.  
 DR P-PSDB; AAY08966.

XX Ashbya gossypii purine biosynthesis enzymes and related genes for  
 PT production of riboflavin  
 XX XX  
 PS Claim 20; Page 35-38; 48pp; German.

XX This invention describes novel enzymes involved in purine biosynthesis  
 CC in Ashbya gossypii. The enzymes described in the invention include  
 CC Phosphoribosylpyrophosphate (PRPP) synthetases KPR1 and KPR2  
 CC glutamine-phosphoribosylpyrophosphate amidotransferase (ADE4)  
 CC IMP-dehydrogenase (GUA1) and a GMP-synthetase (GUA2). The nucleic acid  
 CC sequences encoding the enzymes involved in purine biosynthesis are used  
 CC to genetically modify microorganisms, for the production of riboflavin.  
 CC Riboflavin, also known as vitamin B2, is essential in humans and animals  
 CC and is used to treat inflammation of the mouth and throat mucosal layer  
 CC as well as inflammation in skin disorders. This sequence represents the  
 CC GUA2 gene which encodes a guanosine-monophosphate synthetase.

XX SQ Sequence 2697 BP; 656 A; 631 C; 788 G; 622 T; 0 other;

Query Match 64.8%; Score 16.2; DB 20; Length 2697;  
 Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 acattgtcgaagtcacagaa 22  
 ||||| ||||| ||||| |||||  
 Db 69 AAATCGTGAAGTCAAGTCAAGAA 49

RESULT 22  
 AAT67318/c  
 ID AAT67318 standard; cDNA; 2879 BP.  
 XX AC AAT67318;  
 XX XX

DT 15-DEC-1997 (first entry)  
 XX DE  
 DE Xenopus beta-signalin 4 cDNA.

XX Signalin; Xe-signalin-4; transforming growth factor-beta;  
 KW signal transduction; growth; differentiation; pattern formation;  
 KW hyperproliferative disorder; tumour suppression; spermatogenesis;  
 KW osteogenesis; chondrogenesis; ss.

XX Xenopus sp.  
 XX XX  
 FH Key Location/Qualifiers  
 FT CDS 258..2045  
 FT /\*tag= a

XX WO9722697-A1.

XX 26-JUN-1997.

XX 20-DEC-1996; 96WO-US20745.

XX 20-DEC-1995; 95US-0580031.

XX (HARD ) HARVARD COLLEGE.  
 XX (ONTO-) ONTOGENY INC.

XX Graff JM, Jin P, Melton DA, Woolf TM;

XX WPI; 1997-341688/31.

XX P-PSDB; AAW18097.

XX Vertebrate signalin polypeptide and DNA - used to modulate the  
 PT growth, differentiation or survival of mammalian cells, e.g. for  
 PT treatment of hyperproliferative disorders

XX Claim 40; Page 100-103; 152pp; English.

XX This cDNA clone codes for Xenopus beta-signalin 4 (AAW18097), a  
 CC novel protein that modulates the signal transduction activity of a  
 CC transforming growth factor receptor. It is an example of a novel  
 CC family of vertebrate genes (see also AAT67315-27) whose products (see  
 CC also AAW18094-W18106) have broad apparent involvement in mesoderm  
 CC induction, tumour suppression and the formation and maintenance  
 CC of ordered spatial arrangements of differentiated tissues. The  
 CC clone was isolated from Xenopus embryos by PCR amplification (see  
 CC AAT67328-29). Isolated nucleic acids can be used to express signalin  
 CC polypeptides in prokaryotic or eukaryotic hosts, or in transgenic  
 CC animals, or can be delivered to a cell for use in gene therapy or  
 CC antisense therapy. Mutation or misexpression of a signalin gene  
 CC can be detected using probes, PCR primers or LCR primers based on  
 CC signalin genes. This forms the basis of a claimed diagnostic assay  
 CC to identify cells at risk of a disorder characterised by unwanted  
 CC cell proliferation or differentiation.

XX SQ Sequence 2879 BP; 693 A; 759 C; 721 G; 706 T; 0 other;

Query Match 64.8%; Score 16.2; DB 18; Length 2879;  
 Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gacattgtcgaagtcacagaa 21  
 ||||| ||||| ||||| |||||  
 Db 1716 GACACTGTCGCAAGTCAAGCA 1696

RESULT 23

AAT02799/c  
 ID AAT02799 standard; cDNA to mRNA; 2989 BP.  
 XX  
 AC AAT02799;  
 XX  
 XX  
 DT 26-APR-1996 (first entry)  
 XX  
 DE Human glutamate receptor subunit cDNA (GlurcflopG).  
 XX  
 KW glutamate receptor subunit; human; flip; flop; splice variant;  
 KW identify ligand; treatment; ischaemia; schizophrenia; epilepsy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 73..2739  
 FT /\*tag= a  
 FT /note= "glutamate receptor subunit"  
 XX  
 PN WO9521188-A1.  
 XX  
 XX 10-AUG-1995.  
 PD  
 XX  
 PF 27-JAN-1995; 95WO-EP00290.  
 XX  
 XX 07-FEB-1994; 94DE-4403666.  
 XX  
 XX (BADI ) BASF AG.  
 XX  
 PI Bach A, Hoeger T, Lemaire H, Sterrer S, Ultsch A;  
 XX  
 DR WPI; 1995-283726/37.  
 DR P-PSDB; AAR84916.  
 XX  
 XX DNA sequences coding for glutamate receptor sub-unit variants -  
 PT useful for the identification of functional ligands for these  
 PT receptors and for preparing anti-sense reagents  
 XX  
 PS Claim 1; Page 46-50; 71pp; English.  
 XX  
 CC AAT02795-802 are DNA sequences encoding variants of the human glutamate  
 CC receptor (GlurB) subunits. cDNAs encoding Glur subunits A, B, C, and D,  
 CC were isolated from the brain. There are 2 splice variants, "flip" and  
 CC "flop", for each subunit that are known. It was found that for GlurA,  
 CC GlurB, GlurC and GlurD, the first amino acid of the "flip/flop"-region,  
 CC can be Glycine (G) or Arginine (R), through different RNA splicing.  
 CC For the GlurB subunit, it was found that it is possible to have a  
 CC Glutamine (Q) or an R in the transmembrane domain, before the flip/flop  
 CC region, this gives a possible 8 conformations. Cells transfected with  
 CC the DNA are useful for identifying functional ligands for Glur.  
 CC Glutamate is an important excitatory neurotransmitter in the central  
 CC nervous system and is involved in the pathophysiology of e.g., epilepsy,  
 CC schizophrenia and ischaemia. Glur ligands have the potential to be  
 CC used in the treatment of these conditions.  
 XX  
 SQ Sequence 2989 BP; 868 A; 641 C; 728 G; 752 T; 0 other;  
 XX

Query Match 64.8%; Score 16.2; DB 16; Length 2989;  
 Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 XX  
 QY 3 cattgtcgcaagtcacagaat 23  
 ||||| ||||| |||||  
 Db 2939 CATTTGTCGAGTCTCAGAGT 2919  
 XX  
 RESULT 24  
 AAT02800/c  
 ID AAT02800 standard; cDNA to mRNA; 2989 BP.  
 XX  
 AC AAT02800;  
 XX

DT 26-APR-1996 (first entry)  
 XX  
 DE Human glutamate receptor subunit cDNA (GlurcflopG).  
 XX  
 KW glutamate receptor subunit; human; flip; flop; splice variant;  
 KW identify ligand; treatment; ischaemia; schizophrenia; epilepsy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 73..2739  
 FT /\*tag= a  
 FT /note= "glutamate receptor subunit"  
 XX  
 PN WO9521188-A1.  
 XX  
 XX 10-AUG-1995.  
 PD  
 XX  
 PF 27-JAN-1995; 95WO-EP00290.  
 XX  
 XX 07-FEB-1994; 94DE-4403666.  
 XX  
 XX (BADI ) BASF AG.  
 XX  
 PI Bach A, Hoeger T, Lemaire H, Sterrer S, Ultsch A;  
 XX  
 DR WPI; 1995-283726/37.  
 DR P-PSDB; AAR84917.  
 XX  
 XX DNA sequences coding for glutamate receptor sub-unit variants -  
 PT useful for the identification of functional ligands for these  
 PT receptors and for preparing anti-sense reagents  
 XX  
 PS Claim 1; Page 54-58; 71pp; English.  
 XX  
 CC AAT02795-802 are DNA sequences encoding variants of the human glutamate  
 CC receptor (Glur) subunits. cDNAs encoding Glur subunits A, B, C, and D,  
 CC were isolated from the brain. There are 2 splice variants, "flip" and  
 CC "flop", for each subunit that are known. It was found that for GlurA,  
 CC GlurB, GlurC and GlurD, the first amino acid of the "flip/flop"-region,  
 CC can be Glycine (G) or Arginine (R), through different RNA splicing.  
 CC For the GlurB subunit, it was found that it is possible to have a  
 CC Glutamine (Q) or an R in the transmembrane domain, before the flip/flop  
 CC region, giving a possible 8 conformations. The GlurflopG cDNA  
 CC possesses a base change at position 2377, a glycine codon (GGA) is  
 CC converted to an arginine codon (AGA) (sic). Cells transfected with the  
 CC DNA are useful for identifying functional ligands for Glur. Glutamate  
 CC is an important excitatory neurotransmitter in the central nervous  
 CC system and is involved in the pathophysiology of e.g., epilepsy,  
 CC schizophrenia and ischaemia. Glur ligands have the potential to be  
 CC used in the treatment of these conditions.  
 XX  
 SQ Sequence 2989 BP; 866 A; 641 C; 730 G; 752 T; 0 other;  
 XX

Query Match 64.8%; Score 16.2; DB 16; Length 2989;  
 Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 XX  
 QY 3 cattgtcgcaagtcacagaat 23  
 ||||| ||||| |||||  
 Db 2939 CATTTGTCGAGTCTCAGAGT 2919  
 XX  
 RESULT 25  
 AAQ62694/c  
 ID AAQ62694 standard; cDNA; 3070 BP.  
 XX  
 AC AAQ62694;  
 XX  
 DT 29-JUN-1994 (first entry)  
 XX  
 DE Human Glur3B receptor DNA.

XX GluR3B receptor; glutamate receptor; AMPA-binding receptor;  
KW CNS; central nervous system; plasmid pBS/humGluR3B; ds.  
OS Homo sapiens.

XX Key Location/Qualifiers  
FH 79..2745  
FT /\*tag= a  
FT sig\_peptide 79..144  
FT /\*tag= b  
FT mat\_peptide 145..2745  
FT /\*tag= c

XX EP574257-A.  
XX 15-DEC-1993.  
XX 10-JUN-1993; 93EP-0304500.  
XX 10-JUN-1992; 92US-0896437.  
XX 10-JUN-1992; 92US-0896611.  
XX 10-JUN-1992; 92US-0896612.

XX (ALLX ) ALLELIX BIOPHARMACEUTICALS.  
XX (ELLI ) ELLIOTT C.  
XX (KAMB ) KAMBOJ R.  
XX (NUTT ) NUTT S L.

XX Elliott C, Kamboj R, Nutt SL;  
PI P-PSDB: AAR45143.  
XX WPI: 1993-396903/50.

XX Polynucleotide(s) encoding AMPA-binding human glutamate receptors  
FT - used for providing prods. for use in screening assays or in  
FT detection and/or localisation

XX Claim 1; Page 35-39; 101pp; English.  
XX The GluR3A receptor DNA is contained within plasmid pBS/humGluR3A  
XX (ATCC 75218) and may be expressed by a mammal recombinant cell. The  
XX DNA can be used to obtain products for evaluating the human nervous  
XX system and particularly for assessing potentially therapeutic  
XX interactions between the AMPA-binding human excitatory amino acid  
XX receptors and natural and synthetic ligands. It can also be used  
XX for detecting the presence and/or location of human GluR receptors  
XX e.g. in brain tissue. The DNA was isolated from human brain cDNA  
XX libraries using rat GluR receptor gene probes.

XX Sequence 3070 BP; 899 A; 652 C; 738 G; 780 T; 1 other;  
XX Query Match 64.8%; Score 16.2; DB 14; Length 3070;  
XX Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
XX Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

XX QY 3 cattgtcgcaagtcacagaat 23  
XX ||||| ||||| ||||| |||||  
XX Db 2946 CATTGTCGGAGTCACAGT 2926

XX RESULT 26  
XX AAF21307  
XX ID AAF21307 standard; DNA: 32351 BP.  
XX AC AAF21307;  
XX DT 14-MAR-2001 (first entry)

XX Human low adenosine antisense oligonucleotide related sequence #2874.  
XX Low adenosine antisense oligonucleotide; phosphorothioate; allergy;

KW human; airway disorder; bronchoconstriction; lung inflammation;  
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;  
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;  
KW respiratory obstruction; pulmonary obstruction; impeded respiration;  
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;  
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;  
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;  
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;  
KW cancer; ss.

XX Homo sapiens.  
XX WO200062736-A2.  
XX 26-OCT-2000.  
XX 24-MAR-2000; 2000WO-US08020.  
XX 06-APR-1999; 99US-0127958.  
XX (UYEC-) UNIV EAST CAROLINA.  
XX (NYCE/) NYCE J W.  
XX Nyce JW;

XX WPI: 2000-679539/66.  
XX Low adenosine (A) content antisense oligonucleotides which do not  
XX trigger adenosine receptors during metabolism, useful e.g. for treating  
XX cancers and respiratory obstructions -  
XX Disclosure; Page 1295-1303; 1592pp; English.

XX The present invention describes low adenosine (A) content antisense  
XX oligonucleotides and compositions (I) comprising them. In the antisense  
XX oligonucleotides the A is replaced by a 'Universal' or alternative base.  
XX (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,  
XX immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.  
XX The antisense oligonucleotides and (I) can be used to down-regulate the  
XX expression and or activity of target polypeptides associated with  
XX lung/respiratory disorders and malignancies, such as stimulating and  
XX activating peptide factors and transmitters, transcription factors,  
XX immunoglobulins and antibodies, antibody receptors, cytokines and  
XX chemokines, endogenously produced specific and non-specific enzymes,  
XX binding proteins, adhesion molecules and their receptors, cytokine and  
XX chemokine receptors, adenosine receptors, bradykinin receptors, central  
XX nervous system (CNS) and peripheral nervous and non-nervous system peptide  
XX receptors, CNS and peripheral nervous and non-nervous system peptide  
XX transmitters, defensins, growth factors, vasoactive peptides and  
XX receptors, binding proteins and malignancy associated proteins. The  
XX antisense oligonucleotides may be used in this way to treat disorders  
XX including respiratory obstruction (especially pulmonary obstruction  
XX and/or bronchoconstriction) and/or lung inflammation, allergy(ies)  
XX and/or surfactant hypoproduction which are associated with a disease or  
XX condition selected from pulmonary vasoconstriction, inflammation,  
XX allergies, asthma, impeded respiration, respiratory distress syndrome  
XX (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary  
XX hypertension, emphysema, chronic obstructive pulmonary disease (COPD),  
XX pulmonary transplantation rejection, pulmonary infections, bronchitis,  
XX and/or cancer. AAF18434 to AAF21543 represent human polynucleotide  
XX fragments and antisense oligonucleotides used in the exemplification of  
XX the present invention.

XX Sequence 32351 BP; 8594 A; 7026 C; 7405 G; 9326 T; 0 other;  
XX Query Match 64.8%; Score 16.2; DB 21; Length 32351;  
XX Best Local Similarity 85.7%; Pred. No. 2.1e+02;  
XX Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

XX QY 5 ttgtcgcaagtcacagaatta 25  
XX ||||| ||||| ||||| |||||  
XX Db 10689 ttgccaagtcacagaatta 10709

RESULT 27  
 ID AAA35185  
 XX AAA35185 standard; DNA; 32351 BP.  
 XX AAA35185;  
 XX 28-JUL-2000 (first entry)  
 DT Human adenosine receptor related polynucleotide 2nd SEQ ID NO:59.  
 DE Human:  
 XX Human: adenosine receptor; low adenosine antisense oligonucleotide;  
 KW phosphorothioate; impaired respiration; inflammation; allergy;  
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;  
 KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;  
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;  
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;  
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;  
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX W0200009525-A2.  
 XX  
 XX 24-FEB-2000.  
 XX  
 XX 03-AUG-1999; 99WO-US17712.  
 XX  
 XX 03-AUG-1998; 98US-0095212.  
 PR  
 XX (UYEC-) UNIV EAST CAROLINA.  
 PA  
 XX Nyce JW;  
 XX  
 XX WPI; 2000-205971/18.  
 DR  
 XX  
 XX New antisense oligonucleotides useful for treating e.g. pulmonary  
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,  
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or  
 PT cancers  
 XX  
 PS Disclosure; Page 1212-1219; 1343pp; English.  
 XX  
 CC The present invention describes a new composition comprising an  
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which  
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or  
 CC inflammation. The ON can have antiinflammatory, antiallergic,  
 CC antiasthmatic, cytostatic and analgesic activities. The compositions are  
 CC useful for the treatment of diseases associated with inflammation,  
 CC impaired airways, including lung disease and diseases whose secondary  
 CC effects afflict the lungs of a subject. They can be used for treating  
 CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,  
 CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic  
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive  
 CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,  
 CC carcinomas, and cancers which may metastasize to the lungs, including  
 CC breast and prostate cancer. The reduction of the adenosine content of  
 CC the ONs reduces side effects. The A-containing ONs break down with the  
 CC release of deoxyadenosine which activates adenosine receptors causing  
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the  
 CC nucleotide sequences given in the sequence listing from the present  
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last  
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences  
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680  
 CC (AAA32323 to AAA33992) are specifically claimed ONs from the present  
 CC invention. N.B. Sequences given in the disclosure of the present  
 CC invention do not match up with their corresponding SEQ ID NO: sequences  
 CC given in the sequence listing.  
 XX  
 XX Sequence 32351 BP; 8594 A; 7027 C; 7405 G; 9325 T; 0 other;  
 SQ

Query Match 64.8%; Score 16.2; DB 21; Length 32351;  
 Best Local Similarity 85.7%; Pred. No. 2.1e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Oy 5 ttgtcgcagtcacagaatta 25  
 ||| | ||||| ||||| |||||  
 Db 10689 ttgccaaagtcacagaatta 10709  
 RESULT 28  
 AAF21311  
 ID AAF21311 standard; DNA; 40298 BP.  
 XX  
 XX AAF21311;  
 AC  
 XX 14-MAR-2001 (first entry)  
 DT  
 XX Human low adenosine antisense oligonucleotide related sequence #2878.  
 DE  
 XX Human:  
 KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;  
 KW human; airway disorder; bronchoconstriction; lung inflammation;  
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;  
 KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;  
 KW respiratory obstruction; pulmonary obstruction; impeded respiration;  
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;  
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;  
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;  
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;  
 KW cancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX W0200062736-A2.  
 PN  
 XX 26-OCT-2000.  
 PD  
 XX  
 XX 24-MAR-2000; 2000WO-US08020.  
 PF  
 XX 06-APR-1999; 99US-0127958.  
 PR  
 XX (UYEC-) UNIV EAST CAROLINA.  
 PA  
 XX (NYCE/) NYCE J W.  
 XX  
 XX Nyce JW;  
 PI  
 XX  
 XX WPI; 2000-679539/66.  
 DR  
 XX Low adenosine (A) content antisense oligonucleotides which do not  
 PT trigger adenosine receptors during metabolism, useful e.g. for treating  
 PT cancers and respiratory obstructions -  
 PT  
 PS Disclosure; Page 1305-1315; 1592pp; English.  
 XX  
 CC The present invention describes low adenosine (A) content antisense  
 CC oligonucleotides and compositions (I) comprising them. In the antisense  
 CC oligonucleotides the A is replaced by a 'Universal' or alternative base.  
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,  
 CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.  
 CC The antisense oligonucleotides and (I) can be used to down-regulate the  
 CC expression and or activity of target polypeptides associated with  
 CC lung/respiratory disorders and malignancies, such as stimulating and  
 CC activating peptide factors and transmitters, transcription factors,  
 CC immunoglobulins and antibodies, antibody receptors, cytokines and  
 CC chemokines, endogenously produced specific and non-specific enzymes,  
 CC binding proteins, adhesion molecules and their receptors, cytokine and  
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central  
 CC nervous system (CNS) and peripheral nervous and non-nervous system  
 CC receptors, CNS and peripheral nervous and non-nervous system peptide  
 CC transmitters, defensins, growth factors, vasoactive peptides and  
 CC receptors, binding proteins and malignancy associated proteins. The  
 CC antisense oligonucleotides may be used in this way to treat disorders  
 CC including respiratory obstruction (especially pulmonary obstruction  
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)



CC and/or surfactant hypoproduction which are associated with a disease or  
 CC condition selected from pulmonary vasoconstriction, inflammation,  
 CC allergies, asthma, impeded respiration, respiratory distress syndrome  
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary  
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),  
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,  
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide  
 CC fragments and antisense oligonucleotides used in the exemplification of  
 CC the present invention.

XX Sequence 40298 BP; 10485 A; 9119 C; 9484 G; 11210 T; 0 other;  
 SQ

Query Match 64.8%; Score 16.2; DB 21; Length 40298;  
 Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ttgtcgcaagtcacagaatta 25  
 ||| | |||||  
 Db 11514 ttgcccagaagtcacagaatta 11534

RESULT 29  
 AAA35189  
 ID AAA35189 standard; DNA; 40298 BP.

XX AAA35189;

XX 28-JUL-2000 (first entry)

XX Human adenosine receptor related polynucleotide 2nd SEQ ID NO:63.

XX Human; adenosine receptor; low adenosine antisense oligonucleotide;  
 KW phosphorothioate; impaired respiration; inflammation; allergy;  
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;  
 KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;  
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;  
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;  
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;  
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

XX Homo sapiens.

XX WO200009525-A2.

XX 24-FEB-2000.

XX 03-AUG-1999; 99WO-US17712.

XX 03-AUG-1998; 98US-0095212.

XX (UYEC-) UNIV EAST CAROLINA.

XX Nyce JW;

XX WPI; 2000-205971/18.

XX New antisense oligonucleotides useful for treating e.g. pulmonary  
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,  
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or  
 PT cancers -

XX Disclosure; Page 1221-1231; 1343pp; English.

XX The present invention describes a new composition comprising an  
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which  
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or  
 CC inflammation. The ON can have antiinflammatory, antiallergic,  
 CC antiasthmatic, cytostatic and analgesic activities. The compositions are  
 CC useful for the treatment of diseases associated with inflammation,  
 CC impaired airways, including lung disease and diseases whose secondary  
 CC effects afflict the lungs of a subject. They can be used for treating  
 CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,

CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic  
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive  
 CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,  
 CC carcinomas, and cancers which may metastasise to the lungs, including  
 CC breast and prostate cancer. The reduction of the adenosine content of  
 CC the ONs reduces side effects. The A-containing ONs break down with the  
 CC release of deoxyadenosine which activates adenosine receptors causing  
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the  
 CC nucleotide sequences given in the sequence listing from the present  
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last  
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences  
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680  
 CC (AAA32323 to AAA33992) are specifically claimed ONs from the present  
 CC invention. N.B. Sequences given in the disclosure of the present  
 CC invention do not match up with their corresponding SEQ ID NO: sequences  
 CC given in the sequence listing.

XX Sequence 40298 BP; 10485 A; 9121 C; 9482 G; 11210 T; 0 other;

Query Match 64.8%; Score 16.2; DB 21; Length 40298;  
 Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ttgtcgcaagtcacagaatta 25  
 ||| | |||||  
 Db 11514 ttgcccagaagtcacagaatta 11534

RESULT 30

AAAT67395

ID AAT67395 standard; DNA; 420 BP.

XX AAT67395;

XX 08-JUL-1997 (first entry)

XX H. pylori surface or membrane protein ORF 14726542.aa.

XX Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.

XX Helicobacter pylori.

XX Key Location/Qualifiers

FT CDS 1..393

FT FT /\*tag= a  
 FT FT /note= "no stop codon given"

XX WO9640893-A1.

XX 19-DEC-1996.

XX 06-JUN-1996; 96WO-US09122.

XX 01-APR-1996; 96US-0630405.

XX 07-JUN-1995; 95US-0487032.

XX (ASTR ) ASTRA AB.

XX Berglindh OT, Smith D, Mellgaard BL;

XX WPI; 1997-052306/05.

XX P-PSDB; AAW20155.

XX Helicobacter pylori nucleic acid sequences and related  
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
 PT infection, and to detect Helicobacter

XX Claim 27; Pages 149; 1481pp; English.

XX The present sequence encodes a H. pylori surface or membrane protein

CC



CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
CC production, e.g. in E. coli hosts.

SQ Sequence 477 BP; 186 A; 85 C; 76 G; 130 T; 0 other;

Query Match 64.0%; Score 16; DB 18; Length 477;

Best Local Similarity 79.2%; Pred. No. 1.6e+02;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 acattgcgaagtcacagaatta 25

DB 142 aaatctcgaagtcacagaatta 165

## RESULT 33

AAC36423

ID AAC36423 standard; DNA; 537 BP.

XX

AC AAC36423;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 13747.

XX

KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.

PR

PR 05-MAR-1999; 99US-0123180.

PR

PR 09-MAR-1999; 99US-0123548.

PR

PR 23-MAR-1999; 99US-0125788.

PR

PR 25-MAR-1999; 99US-0126264.

PR

PR 29-MAR-1999; 99US-0126785.

PR

PR 01-APR-1999; 99US-0127462.

PR

PR 06-APR-1999; 99US-0128234.

PR

PR 08-APR-1999; 99US-0128714.

PR

PR 16-APR-1999; 99US-0129845.

PR

PR 19-APR-1999; 99US-0130077.

PR

PR 21-APR-1999; 99US-0130449.

PR

PR 23-APR-1999; 99US-0130510.

PR

PR 28-APR-1999; 99US-0130891.

PR

PR 30-APR-1999; 99US-0131449.

PR

PR 30-APR-1999; 99US-0132048.

PR

PR 04-MAY-1999; 99US-0132407.

PR

PR 05-MAY-1999; 99US-0132485.

PR

PR 06-MAY-1999; 99US-0132486.

PR

PR 07-MAY-1999; 99US-0132487.

PR

PR 11-MAY-1999; 99US-0134256.

PR

PR 14-MAY-1999; 99US-0134218.

PR

PR 14-MAY-1999; 99US-0134219.

PR

PR 14-MAY-1999; 99US-0134221.

PR

PR 14-MAY-1999; 99US-0134370.

PR

PR 18-MAY-1999; 99US-0134768.

PR

PR 19-MAY-1999; 99US-0134941.

PR

PR 20-MAY-1999; 99US-0135124.

PR

PR 21-MAY-1999; 99US-0135353.

PR

PR 24-MAY-1999; 99US-0135629.

PR

PR 25-MAY-1999; 99US-0136021.

PR

PR 27-MAY-1999; 99US-0136392.

PR

PR 28-MAY-1999; 99US-0136782.

PR

PR 01-JUN-1999; 99US-0137222.

PR

PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 21-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 22-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match

64.0%; Score 16; DB 21; Length 537;

Best Local Similarity 79.2%; Pred. NO. 1.6e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gacattgtcgaagtcacagaatt 24  
| | | | | | | | | | | | | | | |  
Db 462 ggcgttgcgaagtcacagaatt 485

## RESULT 34

AAA81657/C

ID AAA81657 standard; DNA; 1157 BP.

XX

AC AAA81657;

XX

DT 04-DEC-2000 (first entry)

XX

DE N. meningitidis partial DNA sequence gnm\_204 SEQ ID NO:204.

XX

XX WO200022430-A2.

PD 20-APR-2000.

XX

PF 08-OCT-1999; 99WO-US23573.

XX

PR 09-OCT-1998; 98US-0103794.

XX

PR 30-APR-1999; 99US-0132068.

XX

(CHIR ) CHIRON CORP.

XX

Frazier CM, Hickey E, Peterson J, Tettelin H, Venter JC;

Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;

Rappuoli R, Pizza M;

WPI; 2000-318079/27.

XX

Isolated nucleotide sequences of *Neisseria meningitidis* which can beused in the diagnosis and treatment of *N. meningitidis* infection andother *Neisseria* infections, for example, *N.gonorrhoea*.

Claim 7; Page 1529; 1760pp; English.

XX

The present invention describes methods of obtaining immunogenic

proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414represent specifically claimed *Neisseria meningitidis* genomic DNA

sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent

*Neisseria* DNA sequences and their corresponding proteins; AAA81254 to

AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the

isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 toAAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF

sequences, which are all used in the exemplification of the present

invention. The nucleic acid sequences, protein sequences, and antibodies

against them, can be used in the manufacture of a composition. The

composition can be used as a medicament (or in the manufacture of a

medicament) for treating, preventing or diagnosing infection due to

*Neisseria* bacteria. For example, some of the identified proteins couldbe components of vaccines against *Meningococcus B*; against all serotypes;and/or against all pathogenic *Neisseria*. Identification of sequences

from the bacterium will also facilitate production of biological probes,

particularly organism-specific probes. Attempts to make efficacious

*Meningococcus B* vaccines have failed mainly due to antigen tolerance.

Multivalent vaccines have also been tried but none have successfully

overcome antigenic variability. The provision of further, complete

sequences may provide an opportunity to identify secreted or surface

exposed proteins that may be presumed targets for the immune system and

which are not antigenically variable or at least more conserved than

other more variable regions.

XX

SQ Sequence 1157 BP; 275 A; 275 C; 248 G; 359 T; 0 other;

Query Match 64.0%; Score 16; DB 21; Length 1157;  
 Best Local Similarity 79.2%; Pred. No. 1.8e+02;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 acattgtcgcaagtcacagaatta 25  
 | ||||| ||||| ||||| |||||  
 Db 86 AATTTGCTCAAGCCACACACTA 63

RESULT 35  
 AAA29208/c  
 ID AAA29208 standard; DNA; 2973 BP.  
 XX  
 AC AAA29208;  
 XX  
 DT 12-SEP-2000 (first entry)  
 XX  
 DE Human topoisomerase III-alpha coding sequence.  
 XX  
 KW Topoisomerase III-alpha; interacting protein; TIII-a IP; RNA helicase;  
 KW DBX1; nucleosome; supercoiling; chromosome segregation; recombination;  
 KW stabilization; cell division; apoptosis; cell cycle regulation;  
 KW cytotoxic; anti-tumour; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 16..2940  
 FT /\*tag= a  
 FT /product= Topoisomerase\_III-alpha

WO200032768-A1.  
 08-JUN-2000.  
 29-NOV-1999; 99WO-FR02952.  
 30-NOV-1998; 98FR-0015081.  
 (AVET ) AVENTIS PHARMA SA.  
 Fournier A, Goulaouic H, Riou J;  
 WPI; 2000-412316/35.  
 P-PSDB; AAY96484.

New nucleic acid encoding ligand for topoisomerase III alpha, useful  
 for inhibiting the enzyme and in drug screening, e.g. for potential  
 anticancer agents

Example 1; Page 34-39; 68pp; French.

This DNA encodes human topoisomerase III-alpha (TIII-a). A novel TIII-a  
 interacting protein (TIII-a IP) has been identified (see AAY96482).  
 TIII-a IP has structural features in common with RNA helicases (e.g.  
 DBX1), which are involved in destruction of the nucleosome structure,  
 supercoiling of DNA, segregation of newly replicated chromosomes, and  
 recombination and stabilization of the genome. Agents that alter  
 interaction between TIII-a IP and TIII-a therefore modulate cell  
 division, replication, transcription, translation, splicing and DNA  
 recombination or repair, so may slow cell growth, block the cell cycle  
 or induce apoptosis. Antibodies and ligands of TIII-a IP are used to  
 prevent, treat or alleviate diseases that involve abnormal regulation of  
 the cell cycle, i.e. they are potential anti-tumour (cytotoxic) agents.

Sequence 2973 BP; 734 A; 793 C; 833 G; 613 T; 0 other;

Query Match 64.0%; Score 16; DB 21; Length 2973;  
 Best Local Similarity 79.2%; Pred. No. 2e+02;

SQ Sequence 1157 BP; 275 A; 275 C; 248 G; 359 T; 0 other;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gacattgtcgcaagtcacagaatt 24  
 | | | | | | | | | | | | | | | | | |  
 Db 2382 GCCACAGTTGCAGTCACAGAATT 2359

RESULT 36  
 AAA13178/c  
 ID AAA13178 standard; DNA; 4812 BP.  
 XX  
 AC AAA13178;  
 XX  
 DT 20-JUL-2000 (first entry)  
 XX  
 DE Vitellogenin encoding gene vit-5 nucleotide sequence.  
 XX  
 KW Vitellogenin; vit-5; primer set; detect; endocrine disturbance;  
 KW toxicity; environmental contamination; ds.  
 XX  
 OS Caenorhabditis elegans.  
 XX  
 PN JP2000069977-A.  
 XX  
 PD 07-MAR-2000.  
 XX  
 PF 03-SEP-1998; 98JP-0249723.  
 XX  
 PR 03-SEP-1998; 98JP-0249723.  
 XX  
 PA (KUMA-) KUMAMOTO KOTAI KENKYUSHO KK.  
 XX  
 DR WPI; 2000-264452/23.  
 XX  
 PT A primer for the amplification of nematode vitellogenin gene - for  
 evaluation of the endocrine disturbing activity of a chemical substance

Claim 3; Page 12-13; 17pp; Japanese.

This sequence represents the Caenorhabditis elegans vit-5 gene, which  
 encodes vitellogenin. The present invention relates to a primer set (see  
 AAA13179-A13188) containing primers which hybridise to different parts  
 of the common regions of the vit-1, vit-2, vit-3, vit-4 and vit-5 genes.  
 The vit-1 to 5 genes encode nematode vitellogenin. The invention also  
 includes a method for the detection of a vitellogenin gene in which a  
 gene encoding nematode vitellogenin is amplified using the primer set. A  
 method for evaluating the endocrine disturbing activity of a chemical  
 substance in which a gene encoding vitellogenin is amplified using the  
 primer set is also included in the invention. The method for evaluation  
 can be used for research on the toxicity of chemical substances and  
 environmental contamination.

Sequence 4812 BP; 1381 A; 1281 C; 966 G; 1184 T; 0 other;

Query Match 64.0%; Score 16; DB 21; Length 4812;  
 Best Local Similarity 79.2%; Pred. No. 2.1e+02;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gacattgtcgcaagtcacagaatt 24  
 | | | | | | | | | | | | | | | | | |  
 Db 822 GACCTTGGCGGAGTCTCAGACTT 799

RESULT 37  
 AAA94634/c  
 ID AAA94634 standard; DNA; 5994 BP.  
 XX  
 AC AAA94634;  
 XX  
 DT 15-JAN-2001 (first entry)  
 XX  
 DE Human TUB 5' region.

```

XX KW Human; TUB; neurosensory defect; retina; retinal dystrophy; ds.
XX OS Homo sapiens.
XX PN US6114502-A.
XX PD 05-SEP-2000.
XX PF 27-FEB-1998; 98US-0032365.
XX PR 22-AUG-1996; 96US-0701380.
XX PR 04-SEP-1996; 96US-0706292.
XX PR 10-APR-1996; 96US-0630592.
XX PR 17-SEP-1996; 96US-0714991.
XX PR 30-APR-1997; 96US-0850218.
XX PR 01-AUG-1997; 97US-0904699.
XX PR 17-SEP-1997; 97US-0932306.
XX PA (AXYS-) AXYS PHARM INC.
XX PI North M, Nishina P, Noben-Trauth K, Naggert J;
XX DR WPI; 2000-586483/55.
XX PT Mammalian proteins expressed in retina and brain, useful for producing
XX PT antibodies and for diagnosing neurosensory defects including cochlear
XX PT degeneration, peripheral retinal degeneration and cone-rod retinal
XX PT dystrophy -
XX PS Disclosure; Columns 51-58; 61pp; English.
XX CC TUB is a member of the neurosensory defect associated gene family, and is
XX CC expressed in the retina. There are 6 alternative 5' ends for the TUB
XX CC transcript, which lead to different amino acid sequences for the
XX CC N-terminus. The present sequence is one such 5' region from human
XX CC TUB. The TUB gene is useful as an immunogen to raise antibodies that
XX CC specifically identify TUB expressing cells and in drug screening assays
XX CC directed at neurosensory defects. The TUB protein can be used for the
XX CC treatment of neurosensory degenerative conditions e.g. retinal
XX CC dystrophies.
XX SQ Sequence 5994 BP; 1352 A; 1505 C; 1663 G; 1474 T; 0 other;

Query Match 64.0%; Score 16; DB 21; Length 5994;
Best Local Similarity 79.2%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 acattgtcgcaagtcacagaatta 25
Db 3110 ACAGTGTCTCAAGTCCAGATATA 3087

RESULT 38
ID AAT96641/c
XX AAT96641 standard; DNA; 5995 BP.
XX AC AAT96641;
XX 27-APR-1998 (first entry)
XX DE Human TUB locus 5' region.
XX KW TUB; TUB Form I; sensory neuron; neurosensory defect;
XX KW cochlear degeneration; hearing loss; deafness; retinal dystrophy;
XX KW retinitis pigmentosa; combined rod cone dystrophy; obesity; human;
XX KW animal model; transgenic animal; therapy; diagnosis; ds.
XX OS Homo sapiens.
XX Key Location/Qualifiers
XX FH misc_difference 2640
XX FT

```

```

FT /*tag= a
FT /note= "n.at position 2640 is given as ':' in
FT the specification"
XX PN WO9738004-A1.
XX PD 16-OCT-1997.
XX PF 10-APR-1997; 97WO-US05903.
XX PR 17-SEP-1996; 96US-0714991.
XX PR 10-APR-1996; 96US-0630592.
XX PR 22-AUG-1996; 96US-0701380.
XX PR 04-SEP-1996; 96US-0706292.
XX PA (JACK-) JACKSON LAB.
XX PA (SEQU-) SEQUANA THERAPEUTICS INC.
XX PI Naggert J, Nishina P, Noben-Trauth K, North M;
XX DR WPI; 1997-512642/47.
XX PT Mammalian TULP protein - used for detecting pre-disposition to
XX PT neuro-sensory defects
XX PS Disclosure; Page 53-54; 89pp; English.
XX CC This sequence comprises the genomic region 5' to the human TUB
XX CC locus 6 cDNA splice variants of TUB have been identified (see
XX CC AAT96682-87) by 5' RACE and are designated as Form 1 through 6. The
XX CC encoded proteins have a common C-terminal sequence (AAW36488) and
XX CC vary in their N-terminal sequences (see AAW36495-97). The human TUB
XX CC locus maps to chromosome 11. The TUB gene is expressed at high
XX CC levels in the brain, eye and testis. TUB is a member of the
XX CC mammalian TULP gene family associated with various defects in
XX CC sensory neurons such as cochlear defects, retinitis pigmentosa and
XX CC combined rod-cone dystrophy. The nucleotide sequences of human
XX CC and mouse TULP cDNAs and genomic regions (see AAT96636-45 and
XX CC AAT96682-87) are provided. These can be used to identify related
XX CC sequences; for production of encoded protein (see AAW36485-97); to
XX CC study associated physiological pathways in vivo and in vitro; to
XX CC breed transgenic or knockout animal models of disease; and to
XX CC modulate (e.g. using antisense sequences) gene activity for
XX CC diagnostic, prophylactic and therapeutic purposes, such as
XX CC treatment of neurosensory defects. The DNA is further used as a
XX CC diagnostic for genetic predisposition to the specific genetically
XX CC linked defect.
XX SQ Sequence 5995 BP; 1352 A; 1506 C; 1662 G; 1474 T; 1 other;

Query Match 64.0%; Score 16; DB 18; Length 5995;
Best Local Similarity 79.2%; Pred. No. 2.2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 acattgtcgcaagtcacagaatta 25
Db 3111 ACAGTGTCTCAAGTCCAGATATA 3088

RESULT 39
ID AAV58288/c
XX AAV58288 standard; DNA; 6688 BP.
XX AC AAV58288;
XX 18-DEC-1998 (first entry)
XX DE S. pyogenes SP-26-46 genomic DNA.
XX KW Probe; diagnosis; infection; detection; ds.
XX OS Streptococcus pyogenes.

```

XX PN WO9842845-A1.  
 XX PD 01-OCT-1998.  
 XX PF 23-MAR-1998; 98WO-JP01288.  
 XX PR 25-MAR-1997; 97JP-0071077.  
 XX PA (FUSO ) FUSO PHARM IND LTD.  
 XX PI Abe K, Keshi H, Matsuhisa A, Ueyama H;  
 XX DR WPI; 1998-532009/45.  
 XX PT New DNA probes, e.g. SP-6-28 or SP-7-44 - useful for, e.g. diagnosis  
 XX PT of Streptococcus pyogenes infection  
 XX PS Claim 2; Page 30-35; 48pp; Japanese.  
 XX CC AAV58284-V58289 are novel genomic DNA sequences which can be used as DNA  
 XX CC probes for the diagnosis of Streptococcus pyogenes infection. These  
 XX CC probes provide for simple and highly specific detection of S. pyogenes  
 XX CC in biological samples such as blood.  
 XX SQ Sequence 6688 BP; 1938 A; 1193 C; 1464 G; 2093 T; 0 other;  
 Query Match 64.0%; Score 16; DB 19; Length 6688;  
 Best Local Similarity 79.2%; Pred. No. 2.2e+02;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 gacattgtcgcaagtcacagaatt 24  
 ||| |||| |||| |||| ||||  
 Db 4228 GACGTTGTGCAATTACGTAATT 4205  
 RESULT 40  
 AAX91990  
 ID AAX91990 standard; DNA; 1230025 BP.  
 XX AC AAX91990;  
 XX DT 13-SEP-1999 (first entry)  
 XX DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.  
 XX KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 XX KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 XX KW vaccine; neutralising epitope; ss.  
 XX OS Chlamydia pneumoniae.  
 XX PN WO9927105-A2.  
 XX PD 03-JUN-1999.  
 XX PF 20-NOV-1998; 98WO-TB01890.  
 XX PR 04-NOV-1998; 98US-0107078.  
 XX PR 21-NOV-1997; 97FR-0014673.  
 XX PA (GEST ) GENSET.  
 XX PI Griffais R;  
 XX DR WPI; 1999-357842/30.  
 XX PT Genome sequence of Chlamydia pneumoniae  
 XX PS Claim 1; Page 291-611; 1912pp; English.  
 XX SQ The present sequence represents the complete genome of Chlamydia

CC pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes  
 CC respiratory disease such as pneumonia and bronchitis and is thought  
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,  
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides  
 CC encoded by the open reading frames of the C. pneumoniae genome (see  
 CC AAY34584-Y35879) can be used in immunogenic compositions as vaccines.  
 CC Vectors containing C. pneumoniae nucleotide sequences can also be  
 CC used as immunogenic compositions, especially where the vector directs  
 CC the expression of a neutralising epitope of C. pneumoniae.  
 XX SQ Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;  
 Query Match 64.0%; Score 16; DB 20; Length 1230025;  
 Best Local Similarity 79.2%; Pred. No. 2.8e+02;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 gacattgtcgcaagtcacagaatt 24  
 ||||| ||||| ||||| |||||  
 Db 196949 gacagttggcaagtcagagatt 196972  
 RESULT 41  
 AAO51028  
 ID AAO51028 standard; cDNA; 60 BP.  
 XX AC AAO51028;  
 XX DT 18-APR-1994 (first entry)  
 XX DE Fragment of the human glutamate receptor gene, HSGluR1.  
 XX KW L-glutamate; receptors; neurological disorders; stroke; epilepsy;  
 XX KW head trauma; neurodegenerative processes; Alzheimer's disease; ss.  
 XX OS Homo sapiens.  
 XX PN EP568384-A.  
 XX PD 03-NOV-1993.  
 XX PF 30-APR-1993; 93EP-0303401.  
 XX PR 01-MAY-1992; 92US-0879688.  
 XX PR 19-MAY-1992; 92US-0885912.  
 XX PA (ELIL ) LILLY & CO ELI.  
 XX PI Burnett JP, Mayne NG, Sharp RL, Snyder YM;  
 XX DR WPI; 1993-346751/44.  
 XX DR P-PSDB; AAR42357.  
 XX PT Human glutamate receptor protein and nucleic acid - used to  
 XX PT develop therapeutic agents for neurological disorders and  
 XX PT neuro-degenerative diseases  
 XX PS Disclosure; Page 5; 42pp; English.  
 XX CC The DNA sequence encodes a fragment of the human glutamate receptor  
 XX CC HSGluR1 from residues 130-189. The full length gene was produced from  
 XX CC a culture of E. coli cells contg. the plasmid pRS103 (obtd. from  
 XX CC American Type Culture Collection, Rockville Maryland). The plasmid  
 XX CC was digested with restriction enzymes EcoRI and KpnI to produce a  
 XX CC 4.2 kb fragment contg. the entire coding sequence of the human  
 XX CC HSGluR1 receptor. The receptor can be used to study interactions  
 XX CC of L-glutamate with its receptor and to develop therapeutic agents  
 XX CC for neurological disorders such as stroke, epilepsy and head trauma  
 XX CC and neurodegenerative disorders such as Alzheimer's disease. Fragments  
 XX CC of the DNA sequence may be used as probes to identify further glutamate  
 XX CC receptor genes.  
 XX CC See also AAO51025-32.

SQ Sequence 60 BP; 13 A; 19 C; 12 G; 16 T; 0 other;

Query Match 63.2%; Score 15.8; DB 14; Length 60;  
Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 cattgtcgcaagtccacaga 21  
| | | | | | | | | | | | | | | |  
Db 5 cttgtcgcaactcacaga 23

## RESULT 42

AAQ76727/c  
ID AAQ76727 standard; DNA; 253 BP.

XX AC

XX AC

XX 23-SEP-1994 (first entry)

XX Human genome fragment.

XX Brain; placenta; bone marrow; genetic analysis; gene mapping;  
KW detection; homology; human; adrenal tissue; ds.  
XX Homo sapiens.

XX WC9401548-A.

XX 20-JAN-1994.

XX 13-JUL-1993; 93WO-GB01467.

XX 13-JUL-1992; 92GB-0014857.

XX (MEDI-) MEDICAL RES COUNCIL.

XX Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;  
PI Sibson DR, Starkey M;

XX WPI; 1994-035056/04.

XX New nucleic acid fragment encoding gene products - can be used  
FT for genetic analysis and mapping  
XX Claim 1; Page 176; 616pp; English.

XX Human nucleic acid fragments, isolated from brain, adrenal tissue,  
CC the placenta or bone marrow comprise any of: (A) a sequence  
CC selected from (AAQ76401-Q77613), (B) an allelic variation of a  
CC sequence as described in (A), or (C) a sequence complementary  
CC to (A) or (B).  
CC Preferred sequences exhibit no more than 90% homology to a human  
CC sequence known per se.

SQ Sequence 253 BP; 71 A; 74 C; 53 G; 54 T; 1 other;

Query Match 63.2%; Score 15.8; DB 15; Length 253;  
Best Local Similarity 89.5%; Pred. No. 1.8e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 acattgtcgcaagtccacag 20  
| | | | | | | | | | | | | | | |  
Db 109 ACATTGTCGAGAGTCCACAG 91

## RESULT 43

AAC41591/c  
ID AAC41591 standard; DNA; 628 BP.

XX AC

XX AAC41591;

XX

DT 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 32428.  
DE Hybridisation assay; genetic mapping; gene expression control;  
XX protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX Arabidopsis thaliana.  
OS Arabidopsis thaliana.  
XX EPI033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.



PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145219.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.

PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 63.2%; Score 15.8; DB 21; Length 628;  
Best Local Similarity 89.5%; Pred. No. 2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 cattgtcgcaagtccacaga 21  
| | | | | | | | | | | | | | | | | |  
Db 533 CTTTGTCTCCCAAGTCACAGA 515

RESULT 44  
AAC39882/c  
ID AAC39882 standard; DNA; 629 BP.  
XX  
AC AAC39882;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 26248.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.

XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX XX  
XX PD 06-SEP-2000.  
XX PF  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
XX 09-MAR-1999; 99US-0123548.  
XX 23-MAR-1999; 99US-0125788.  
XX 25-MAR-1999; 99US-0126264.  
XX 29-MAR-1999; 99US-0126785.  
XX 01-APR-1999; 99US-0127462.  
XX 06-APR-1999; 99US-0128234.  
XX 08-APR-1999; 99US-0128714.  
XX 16-APR-1999; 99US-0129845.  
XX 19-APR-1999; 99US-0130077.  
XX 21-APR-1999; 99US-0130449.  
XX 23-APR-1999; 99US-0130510.  
XX 28-APR-1999; 99US-0130891.  
XX 30-APR-1999; 99US-0131449.  
XX 30-APR-1999; 99US-0132048.  
XX 04-MAY-1999; 99US-0132407.  
XX 05-MAY-1999; 99US-0132484.  
XX 06-MAY-1999; 99US-0132486.  
XX 07-MAY-1999; 99US-0132487.  
XX 11-MAY-1999; 99US-0132863.  
XX 14-MAY-1999; 99US-0134256.  
XX 14-MAY-1999; 99US-0134218.  
XX 14-MAY-1999; 99US-0134219.  
XX 14-MAY-1999; 99US-0134221.  
XX 18-MAY-1999; 99US-0134370.  
XX 19-MAY-1999; 99US-0134768.  
XX 20-MAY-1999; 99US-0134941.  
XX 21-MAY-1999; 99US-0135124.  
XX 24-MAY-1999; 99US-0135353.  
XX 25-MAY-1999; 99US-0135629.  
XX 27-MAY-1999; 99US-0136021.  
XX 28-MAY-1999; 99US-0136392.  
XX 01-JUN-1999; 99US-0136782.  
XX 03-JUN-1999; 99US-0137222.  
XX 04-JUN-1999; 99US-0137528.  
XX 07-JUN-1999; 99US-0137502.  
XX 08-JUN-1999; 99US-0137724.  
XX 10-JUN-1999; 99US-0138094.  
XX 10-JUN-1999; 99US-0138847.  
XX 16-JUN-1999; 99US-0139119.  
XX 16-JUN-1999; 99US-0139452.  
XX 17-JUN-1999; 99US-0139453.  
XX 18-JUN-1999; 99US-0139492.  
XX 18-JUN-1999; 99US-0139454.  
XX 18-JUN-1999; 99US-0139455.  
XX 18-JUN-1999; 99US-0139456.  
XX 18-JUN-1999; 99US-0139457.  
XX 18-JUN-1999; 99US-0139458.  
XX 18-JUN-1999; 99US-0139459.  
XX 18-JUN-1999; 99US-0139460.  
XX 18-JUN-1999; 99US-0139461.  
XX 18-JUN-1999; 99US-0139462.  
XX 18-JUN-1999; 99US-0139463.  
XX 18-JUN-1999; 99US-0139750.  
XX 21-JUN-1999; 99US-0139763.  
XX 22-JUN-1999; 99US-0139817.  
XX 23-JUN-1999; 99US-0139899.  
XX 23-JUN-1999; 99US-0140353.  
XX 23-JUN-1999; 99US-0140354.  
XX 24-JUN-1999; 99US-0140895.  
XX 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.

```

PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 63.2%; Score 15.8; DB 21; Length 629;
Best Local Similarity 89.5%; Pred. No. 2e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 2;

QY 3 cattgtcgcaagtcacaga 21
Db 532 CTTGTCTCCCAAGTCACAGA 514

RESULT 45
AAF08144/C
ID AAF08144 standard; cDNA; 837 BP.
XX
AC AAF08144;
XX
DT 13-MAR-2001 (first entry)
XX
DE Fusarium venenatum EST SEQ ID NO:667.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Fusarium venenatum.
XX
PN WO200056762-A2.
XX

```

---

```

PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO ) NOVO NORDISK BIOTECH INC.
PA (NOVO ) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 86; Page 639; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Trichoderma reesei; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 837 BP; 202 A; 197 C; 233 G; 204 T; 1 other;

Query Match 63.2%; Score 15.8; DB 21; Length 837;
Best Local Similarity 89.5%; Pred. No. 2.1e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 2;

QY 3 cattgtcgcaagtcacaga 21
Db 775 CATTCTCGCAAGTCACCGA 757

Search completed: October 9, 2001, 11:35:54
Job time: 1669 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:08:05 ; Search time 221.13 seconds  
(without alignments)  
21.403 Million cell updates/sec

Title: US-09-396-196F-1  
Perfect score: 25  
Sequence: 1 gacatgtcgcaagtcacagaatta 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PC10S\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appli
2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appli
3	25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appli
4	25	100.0	5872	3	US-08-411-768B-6	Sequence 6, Appli
5	17.8	71.2	3083	1	US-07-718-575-5	Sequence 5, Appli
6	17.8	71.2	3083	1	US-08-481-206-5	Sequence 5, Appli
7	17.8	71.2	3083	2	US-08-486-269A-5	Sequence 5, Appli
8	17	68.0	17	2	US-07-944-259-2	Sequence 2, Appli
9	16.6	66.4	1374	1	US-08-123-761A-2	Sequence 2, Appli
10	16.6	66.4	5198	1	US-08-123-761A-1	Sequence 1, Appli
11	16.2	64.8	1771	2	US-09-166-203-48	Sequence 48, Appli
12	16.2	64.8	2989	1	US-08-687-379-9	Sequence 9, Appli
13	16.2	64.8	2989	1	US-08-687-379-11	Sequence 11, Appli
14	16	64.0	3350	3	US-08-617-860B-3	Sequence 3, Appli
15	16	64.0	5994	3	US-09-032-365A-11	Sequence 11, Appli
16	16	64.0	6688	4	US-09-381-862-5	Sequence 5, Appli
17	15.8	63.2	1188	1	US-08-781-562-2	Sequence 2, Appli
18	15.8	63.2	2946	1	US-08-687-379-1	Sequence 1, Appli
19	15.8	63.2	2946	1	US-08-687-379-3	Sequence 3, Appli
20	15.8	63.2	3220	1	US-08-254-573-1	Sequence 1, Appli
21	15.6	62.4	350	5	PCT-US95-09715-6	Sequence 6, Appli
22	15.6	62.4	350	5	PCT-US95-09715-6	Sequence 6, Appli
23	15.6	62.4	1002	3	US-08-938-546-1	Sequence 1, Appli
24	15.6	62.4	1002	3	US-09-340-812-1	Sequence 1, Appli
25	15.6	62.4	1020	3	US-08-938-546-3	Sequence 3, Appli
26	15.6	62.4	1020	3	US-09-340-812-3	Sequence 3, Appli
27	15.6	62.4	3360	3	US-08-938-546-5	Sequence 5, Appli

28	15.6	62.4	3360	3	US-09-340-812-5	Sequence 5, Appli
29	15.4	61.6	946	3	US-08-916-443A-15	Sequence 15, Appli
30	15.4	61.6	1116	3	US-08-916-443A-16	Sequence 16, Appli
31	15.4	61.6	1592	2	US-08-252-493C-1	Sequence 1, Appli
32	15.4	61.6	1592	3	US-09-276-197-1	Sequence 1, Appli
33	15.4	61.6	2352	2	US-08-889-909A-21	Sequence 21, Appli
34	15.4	61.6	5894	3	US-08-665-259-24	Sequence 24, Appli
35	15.4	61.6	5894	3	US-08-762-500-24	Sequence 24, Appli
36	15.4	61.6	6525	3	US-08-762-500-74	Sequence 74, Appli
37	15	60.0	530	4	US-08-975-762-7	Sequence 7, Appli
38	15	60.0	530	4	US-08-821-324-7	Sequence 7, Appli
39	15	60.0	1749	1	US-08-116-098-1	Sequence 1, Appli
40	15	60.0	6012	1	US-08-375-709-14	Sequence 14, Appli
41	15	60.0	6012	1	US-08-752-929-14	Sequence 14, Appli
42	15	60.0	8010	5	PCT-US95-11859-2	Sequence 2, Appli
43	15	60.0	37895	1	US-08-375-709-1	Sequence 1, Appli
44	15	60.0	37895	1	US-08-752-929-1	Sequence 1, Appli
45	15	60.0	37895	3	US-09-090-793-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-401-068-7  
; Sequence 7, Application US/08401068  
; Patent No. 5859335  
; GENERAL INFORMATION:  
; APPLICANT: Patton, David  
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30B  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/401,068  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/351,970  
; FILING DATE: 08-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Elmer, James Scott  
; REGISTRATION NUMBER: 36,129  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8614  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1041 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1038  
; IDENTIFICATION METHOD: experimental  
; OTHER INFORMATION: /product= "biotin synthase"  
; OTHER INFORMATION: /evidence= EXPERIMENTAL  
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;  
 Best Local Similarity 100.0%; Pred. No. 0.0019;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaatta 25  
 |||  
 Db 21 GACATTGTGCGAAGTCACAGAATTA 45

## RESULT 2

US-08-846-338-7  
 ; Sequence 7, Application US/08846338  
 ; Patent No. 5869719

; GENERAL INFORMATION:  
 ; APPLICANT: Patton, David  
 ; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content  
 ; NUMBER OF SEQUENCES: 20  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: No. 5869719artis Corporation  
 ; STREET: 520 White Plains Road, P.O. Box 2005  
 ; CITY: Tarrytown  
 ; STATE: NY  
 ; COUNTRY: USA  
 ; ZIP: 10591

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30B  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/846,338  
 ; FILING DATE:

; CLASSIFICATION: 800  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Meigs, J. Timothy  
 ; REGISTRATION NUMBER: 38, 241  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 919-541-8587  
 ; TELEFAX: 919-541-8689

; INFORMATION FOR SEQ ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1041 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single

; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; HYPOTHETICAL: NO  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..1038  
 ; IDENTIFICATION METHOD: experimental  
 ; OTHER INFORMATION: /product= "biotin synthase"  
 ; OTHER INFORMATION: /evidence= EXPERIMENTAL  
 ; US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;  
 Best Local Similarity 100.0%; Pred. No. 0.0019;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaatta 25  
 |||  
 Db 21 GACATTGTGCGAAGTCACAGAATTA 45

## RESULT 3

US-08-411-768B-1  
 ; Sequence 1, Application US/08411768B  
 ; Patent No. 6083712

; GENERAL INFORMATION:  
 ; APPLICANT: Olwen Birch  
 ; APPLICANT: Johann Brass  
 ; APPLICANT: Martin Fuhrmann

; APPLICANT: Nicholas Shaw  
 ; TITLE OF INVENTION: Biotechnological Method  
 ; TITLE OF INVENTION: of Producing Biotin  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
 ; STREET: 30 Rockefeller Plaza  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10112

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: WordPerfect  
 ; SOFTWARE: Version 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/411,768B  
 ; FILING DATE: 31-March-95  
 ; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: CH 3124/92  
 ; FILING DATE: 02-OCT-1992  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: CH 2134/93  
 ; FILING DATE: 15-JUL-1993

; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 5872 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double

; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:

; ORGANISM: Escherichia coli  
 ; STRAIN: DSM498  
 ; IMMEDIATE SOURCE:  
 ; CLONE: pBO30A-15/9  
 ; FEATURE:

; NAME/KEY: CDS  
 ; LOCATION: 117..1157  
 ; IDENTIFICATION METHOD: experimental  
 ; OTHER INFORMATION: /codon\_start= 117  
 ; OTHER INFORMATION: /product= "Biotin synthase"  
 ; OTHER INFORMATION: /evidence= EXPERIMENTAL  
 ; OTHER INFORMATION: /gene= "bioB"  
 ; OTHER INFORMATION: /number= 1  
 ; FEATURE:

; NAME/KEY: CDS  
 ; LOCATION: 2295..3050  
 ; OTHER INFORMATION: /codon\_start= 2295  
 ; OTHER INFORMATION: /function= "involved in pimeloyl-CoA synthesis"  
 ; OTHER INFORMATION: /product= "protein"  
 ; OTHER INFORMATION: /gene= "bioC"  
 ; OTHER INFORMATION: /number= 3  
 ; FEATURE:

; NAME/KEY: CDS  
 ; LOCATION: 3750..5039  
 ; IDENTIFICATION METHOD: experimental  
 ; OTHER INFORMATION: /codon\_start= 3750  
 ; OTHER INFORMATION: /PC\_number= 2.6.1.62  
 ; OTHER INFORMATION: /product= "DAPA synthase"  
 ; OTHER INFORMATION: /evidence= EXPERIMENTAL  
 ; OTHER INFORMATION: /gene= "bioA"  
 ; OTHER INFORMATION: /number= 5  
 ; OTHER INFORMATION: /standard\_name= "S-Adenosyl-L-methionine:8-amino-7-oxononanoate aminotransf."  
 ; OTHER INFORMATION: /number= 5  
 ; FEATURE:

; NAME/KEY: CDS  
 ; LOCATION: 117..1157  
 ; IDENTIFICATION METHOD: experimental  
 ; OTHER INFORMATION: /codon\_start= 117  
 ; OTHER INFORMATION: /product= "Biotin synthase"  
 ; OTHER INFORMATION: /evidence= EXPERIMENTAL  
 ; OTHER INFORMATION: /gene= "bioB"  
 ; OTHER INFORMATION: /number= 1  
 ; FEATURE:

; NAME/KEY: CDS  
 ; LOCATION: 2295..3050  
 ; OTHER INFORMATION: /codon\_start= 2295  
 ; OTHER INFORMATION: /function= "involved in pimeloyl-CoA synthesis"  
 ; OTHER INFORMATION: /product= "protein"  
 ; OTHER INFORMATION: /gene= "bioC"  
 ; OTHER INFORMATION: /number= 3  
 ; FEATURE:

; NAME/KEY: CDS  
 ; LOCATION: 3750..5039  
 ; IDENTIFICATION METHOD: experimental  
 ; OTHER INFORMATION: /codon\_start= 3750  
 ; OTHER INFORMATION: /PC\_number= 2.6.1.62  
 ; OTHER INFORMATION: /product= "DAPA synthase"  
 ; OTHER INFORMATION: /evidence= EXPERIMENTAL  
 ; OTHER INFORMATION: /gene= "bioA"  
 ; OTHER INFORMATION: /number= 5  
 ; OTHER INFORMATION: /standard\_name= "S-Adenosyl-L-methionine:8-amino-7-oxononanoate aminotransf."  
 ; OTHER INFORMATION: /number= 5  
 ; FEATURE:

; NAME/KEY: CDS  
 ; LOCATION: 117..1157  
 ; IDENTIFICATION METHOD: experimental  
 ; OTHER INFORMATION: /codon\_start= 117  
 ; OTHER INFORMATION: /product= "Biotin synthase"  
 ; OTHER INFORMATION: /evidence= EXPERIMENTAL  
 ; OTHER INFORMATION: /gene= "bioB"  
 ; OTHER INFORMATION: /number= 1  
 ; FEATURE:

; NAME/KEY: CDS  
 ; LOCATION: 2295..3050  
 ; OTHER INFORMATION: /codon\_start= 2295  
 ; OTHER INFORMATION: /function= "involved in pimeloyl-CoA synthesis"  
 ; OTHER INFORMATION: /product= "protein"  
 ; OTHER INFORMATION: /gene= "bioC"  
 ; OTHER INFORMATION: /number= 3  
 ; FEATURE:

; NAME/KEY: CDS  
 ; LOCATION: 3750..5039  
 ; IDENTIFICATION METHOD: experimental  
 ; OTHER INFORMATION: /codon\_start= 3750  
 ; OTHER INFORMATION: /PC\_number= 2.6.1.62  
 ; OTHER INFORMATION: /product= "DAPA synthase"  
 ; OTHER INFORMATION: /evidence= EXPERIMENTAL  
 ; OTHER INFORMATION: /gene= "bioA"  
 ; OTHER INFORMATION: /number= 5  
 ; OTHER INFORMATION: /standard\_name= "S-Adenosyl-L-methionine:8-amino-7-oxononanoate aminotransf."  
 ; OTHER INFORMATION: /number= 5  
 ; FEATURE:

; NAME/KEY: CDS

```

; LOCATION: 5098..5574
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 5098
; OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
; OTHER INFORMATION: /product= "protein"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ORF1"
; OTHER INFORMATION: /number= 6
; FEATURE:
; NAME/KEY: -10_signal
; LOCATION: 45..49
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "promoter ptac"
; FEATURE:
; NAME/KEY: -35_signal
; LOCATION: 23..28
; OTHER INFORMATION: /standard_name= "promoter ptac"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 105..119
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "bioB RBS no.9"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 2284..2297
; OTHER INFORMATION: /standard_name= "bioC RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 3742..3752
; OTHER INFORMATION: /standard_name= "bioA RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 5088..5100
; OTHER INFORMATION: /standard_name= "ORF1 RBS"
; FEATURE:
; NAME/KEY: terminator
; LOCATION: 5583..5644
; OTHER INFORMATION: /standard_name= "rho-independent
; OTHER INFORMATION: transcriptional terminator"
; FEATURE:
; NAME/KEY: stem_loop
; LOCATION: 5583..5605
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 1..96
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /function= "promoter ptac"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: WO 87/01391 B1
; FILING DATE: 26-AUG-1986
; PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaatta 25
Db 137 GACATTGTCGCAAGTCACAGAATTA 161

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass

```

```

; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411,768B
; FILING DATE: 31-March-95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 3124/92
; FILING DATE: 02-Oct-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 2134/93
; FILING DATE: 15-JUL-1993
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5872 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: DSM498
; IMMEDIATE SOURCE:
; CLONE: pB030A15-9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1154..2308
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 1154
; OTHER INFORMATION: /EC_number= 2.3.1.47
; OTHER INFORMATION: /product= "KAPA synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "biof"
; OTHER INFORMATION: /number= 2
; OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3043..3753
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 3043
; OTHER INFORMATION: /EC_number= 6.3.3.3
; OTHER INFORMATION: /product= "DTB synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "bioD"
; OTHER INFORMATION: /number= 4
; OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 1141..1156
; OTHER INFORMATION: /standard_name= "bioF RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 3030..3045
; OTHER INFORMATION: /standard_name= "bioD RBS"
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: WO 87/01391 B1

```

; FILING DATE: 26-AUG-1986  
; PUBLICATION DATE: 07-APR-1993  
US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;  
Best Local Similarity 100.0%; Pred. No. 0.0025;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgcgaagtcacagaatta 25  
|||||  
DB 137 GACATTGCGCAAGTCACAGAATTA 161

RESULT 5  
US-07-718-575-5/c  
; Sequence 5, Application US/07718575  
; Patent No. 520257  
; GENERAL INFORMATION:  
; APPLICANT: Heinemann Ph.D., Stephen F.  
; APPLICANT: Boulter Ph.D., James R.  
; APPLICANT: Hollmann Ph.D., Michael MNM  
; APPLICANT: Bettler Ph.D., Bernhard MNM  
; APPLICANT: Jensen Ph.D., Jan E.  
; TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
; STREET: 444 So. Flower St., Suite 2000  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: United States  
; ZIP: 90071-2921

COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07718,575  
FILING DATE: 19910813  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Reiter Ph.D., Stephen E.  
REGISTRATION NUMBER: 31192  
REFERENCE/DOCKET NUMBER: P31 8962  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
TELEX: 9103330318

INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3083 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
IMMEDIATE SOURCE:  
CLONE: GLUR3  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 167..2833

US-07-718-575-5

Query Match 71.2%; Score 17.8; DB 1; Length 3083;  
Best Local Similarity 90.5%; Pred. No. 6.8;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 cattgtcgaagtcacagaat 23  
|||||  
DB 3034 CATTGCGCAAGTCACAGT 3014

RESULT 6  
US-08-481-206-5/c  
; Sequence 5, Application US/08481206  
; Patent No. 5739291  
; GENERAL INFORMATION:  
; APPLICANT: Heinemann Ph.D., Stephen F.  
; APPLICANT: Boulter Ph.D., James R.  
; APPLICANT: Hollmann Ph.D., Michael MNM  
; APPLICANT: Bettler Ph.D., Bernhard MNM  
; APPLICANT: Jensen Ph.D., Jan E.  
; TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
; STREET: 444 So. Flower St., Suite 2000  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: United States  
; ZIP: 90071-2921  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/481,206  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/013,767  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Reiter Ph.D., Stephen E.  
REGISTRATION NUMBER: 31192  
REFERENCE/DOCKET NUMBER: P31 8962  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
TELEX: 9103330318  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3083 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
IMMEDIATE SOURCE:  
CLONE: GLUR3  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 167..2833  
US-08-481-206-5

Query Match 71.2%; Score 17.8; DB 1; Length 3083;  
Best Local Similarity 90.5%; Pred. No. 6.8;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 cattgtcgaagtcacagaat 23  
|||||  
DB 3034 CATTGCGCAAGTCACAGT 3014

RESULT 7  
US-08-486-269A-5/c  
; Sequence 5, Application US/08486269A  
; Patent No. 5945509  
; GENERAL INFORMATION:  
; APPLICANT: Heinemann, Stephen F.  
; APPLICANT: Boulter, James R.



```

; APPLICANT: Hollmann, Michael
; APPLICANT: Bettler, Bernhard
; APPLICANT: Jensen, Jan E.
; TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS
; TITLE OF INVENTION: AND METHODS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows DEMONSTRATION Version 2.0D
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,269A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/013,767
; FILING DATE: 04-FEB-1993
; APPLICATION NUMBER: 07/718,575
; FILING DATE: 21-JUN-1991
; APPLICATION NUMBER: PCT/US90/06153
; FILING DATE: 25-OCT-1990
; APPLICATION NUMBER: 07/428,116
; FILING DATE: 27-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9986
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-677-1409
; TELEFAX: 619-677-1465
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; CLONE: GluR3
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 167...2830
; OTHER INFORMATION:
; US-08-486-269A-5

```

```

Query Match 71.2%; Score 17.8; DB 2; Length 3083;
Best Local Similarity 90.5%; Pred. No. 6.8;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 3 cattgtcgcaagtcacagaat 23
| | | | | | | | | | | | | | | |
Db 3034 CATTGCGCAAGTCTCAGAGT 3014

```

```

RESULT 8
US-07-944-259-2/c
; Sequence 2, Application US/07944259
; Patent No. 5885792
; GENERAL INFORMATION:
; APPLICANT: Ifuku, Ohji
; APPLICANT: Haze, Shinitiro
; APPLICANT: Kishimoto, Jiro
; APPLICANT: Nakahama, Kazuo

```

```

; TITLE OF INVENTION: BIOTIN OPERON
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wegner, Cantor, Mueller & Player
; STREET: 1233 20th Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036-8218
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/944,259
; FILING DATE: 19920914
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Cantor, Herbert I.
; REGISTRATION NUMBER: 24,392
; REFERENCE/DOCKET NUMBER: P-450-23557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-887-0400
; TELEFAX: 202-835-0605
; TELEX: 440706
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; US-07-944-259-2

Query Match 68.0%; Score 17; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gtcgcaagtcacagaat 23
| | | | | | | | | | | | | | |
Db 17 GTCGCAAGTCACAGAAT 1

RESULT 9
US-08-123-761A-2/c
; Sequence 2, Application US/08123761A
; Patent No. 5589611
; GENERAL INFORMATION:
; APPLICANT: Briggs, Steven P.
; APPLICANT: Johal, Gurmukh S.
; TITLE OF INVENTION: A DISEASE RESISTANCE GENE FROM MAIZE AS
; TITLE OF INVENTION: A SELECTABLE MARKER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
; STREET: 700 Capital Square, 400 Locust Street
; CITY: Des Moines
; STATE: Iowa
; COUNTRY: USA
; ZIP: 50309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/123,761A
; FILING DATE: 17-SEP-1993
; CLASSIFICATION: 800

```

ATTORNEY/AGENT INFORMATION:  
NAME: Roth, Michael J.  
REGISTRATION NUMBER: 29,342  
REFERENCE/DOCKET NUMBER: 212-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (515) 248-4800  
TELEFAX: (515) 248-4844  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1374 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-123-761A-2

Query Match 66.4%; Score 16.6; DB 1; Length 1374;  
Best Local Similarity 82.6%; Pred. No. 23;  
Matches 19; Conservative 0; Mismatches 4; Indels 0

Qy      2    acattgtcgcaagtcacagaatt    24  
         | | | | | | | | | |  
Db     1329    AAATTATCGCAAGTCATAAAATT    1307

```

RESULT 10.
US-08-123-761A-1/c
; Sequence 1, Application US/08123761A
; Patent No. 5589611
; GENERAL INFORMATION:
; APPLICANT: Briggs, Steven P.
; APPLICANT: Johal, Gurmukh S.
; TITLE OF INVENTION: A DISEASE RESISTANCE GENE FROM MAIZE AS
; TITLE OF INVENTION: A SELECTABLE MARKER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
; STREET: 700 Capital Square, 400 Locust Street
; CITY: Des Moines
; STATE: Iowa
; COUNTRY: USA
; ZIP: 50309

```

Query Match 66.4%; Score 16.6; DB 1; Length 5198;  
Best Local Similarity 82.6%; Pred. No. 28;  
Matches 19; Conservative 0; Mismatches 4; Indels 0

Qy            2 acattgtcgcaagtcacagaatt 24  
             | ||| ||||| ||||| | ||||  
Db           4891 AAATTATCGCAAGTCATAAAATT 4869

```

RESULT 11
US-09-166-203-48
; Sequence 48, Application US/09166203A
; Patent No. 5968826
; GENERAL INFORMATION:
; APPLICANT: Bennett, C. Frank
; APPLICANT: Condon, Tom P.
; APPLICANT: Coswert, Lex M.
; TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
; FILE REFERENCE: ISPH-0323
; CURRENT APPLICATION NUMBER: US/09/166,203A
; CURRENT FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 60
; SEQ ID NO 48
; LENGTH: 1771
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1193)..(1387)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1709)..(1771)
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: I20788 Genbank
; DATABASE ENTRY DATE: 1996-04-18
US-09-166-203-48

```

Query Match 64.8%; Score 16.2; DB 2; Length 1771;  
Best Local Similarity 85.7%; Pred. No. 37;  
Matches 18; Conservative 0; Mismatches 3; Indels 0

Qy      4    attgtcgcaagtccagaatt    24  
         |    ||| ||||| |||||  
Db     306 agtgaaggcaagtcacagaatt    326

```

RESULT 12
US-08-687-379-g/c
: Sequence 9, Application US/08687379
: Patent No. 5756697
:
: GENERAL INFORMATION:
: APPLICANT: Hoeger, Thomas
: APPLICANT: Ullsch, Andreas
: APPLICANT: Bach, Alfred
: APPLICANT: Sterrer, Sylvia
: APPLICANT: Lemaire, Hans-Georg
:
: TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
:
: TITLE OF INVENTION: Preparation and Their Use
:
: NUMBER OF SEQUENCES: 26
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Keil & Weinkauff
: STREET: 1101 Connecticut Avenue
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20036
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
: COMPUTER: IBM AT-compatible, 80286 processor
: OPERATING SYSTEM: MS-DOS version 6.0
: SOFTWARE: Wordperfect version 5.1
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/687,379
: FILING DATE: 05-AUG-1996
: CLASSIFICATION: 435
: INFORMATION FOR SEQ ID NO: 9:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 2989 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; DEVELOPMENTAL STAGE: adult
; TISSUE TYPE: brain
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 73..2736
US-08-687-379-9

```

```

Query Match      64.8%; Score 16.2; DB 1; Length 2989;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 3 cattgtcgcaagtcacagaatt 23
||||| ||||| ||||| |||||
DB 2939 CATTGCGGAGTCTCAGAGT 2919

```

## RESULT 13

```

US-08-687-379-11/c
; Sequence 11, Application US/08687379
; Patent No. 5756697
; GENERAL INFORMATION:

```

```

; APPLICANT: Hoeger, Thomas
; APPLICANT: Ultsch, Andreas
; APPLICANT: Bach, Alfred
; APPLICANT: Steirer, Sylvia
; APPLICANT: Lemaire, Hans-Georg
; TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
; PREPARATION AND THEIR USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kell & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,379
; FILING DATE: 05-AUG-1996
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2989 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; DEVELOPMENTAL STAGE: adult
; TISSUE TYPE: brain
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 73..2736
US-08-687-379-11

```

```

Query Match      64.8%; Score 16.2; DB 1; Length 2989;
Best Local Similarity 85.7%; Pred. No. 40;

```

```

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 cattgtcgcaagtcacagaatt 23
||||| ||||| ||||| |||||
DB 2939 CATTGCGGAGTCTCAGAGT 2919

```

## RESULT 14

```

US-08-617-860B-3/c
; Sequence 3, Application US/08617860B
; Patent No. 6133506
; GENERAL INFORMATION:

```

```

; APPLICANT: Typfer, R., Bautor, J., Bothmann, H., Filsak, E.,
; APPLICANT: Hvrcke-Grandpierre, C., Klein, B., Martini, N.,
; APPLICANT: Mller, A., Schulte, W., Voetz, M., Walek, J.,
; APPLICANT: Schell, J.
; TITLE OF INVENTION: Promoters
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steinberg, Raskin & Davidson, P.C.
; STREET: 1140 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,860B
; FILING DATE: 01-MAR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP94/02950
; FILING DATE: 05-SEP-1994
; APPLICATION NUMBER: DE P4329951.2
; FILING DATE: 04-SEP-1993
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3350 Base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: Double stranded
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Brassica napus
; IMMEDIATE SOURCE:
; LIBRARY: genomic Lambda FIX II
; CLONE: BnACCaseg10
; FEATURE:
; NAME/KEY: Startcodon
; LOCATION: 2611..2613
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(2611..2908, 3001..3341)
US-08-617-860B-3

```

```

Query Match      64.0%; Score 16; DB 3; Length 3350;
Best Local Similarity 79.2%; Pred. No. 51;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 gacattgtcgcaagtcacagaatt 24
||||| ||||| ||||| |||||
DB 443 GACCTTGTGCGGAAATGAGAGAATT 420

```

## RESULT 15

```

US-09-032-365A-11/c
; Sequence 11, Application US/09032365A

```

RESULT 16  
US-09-381-862-5/c  
; Sequence 5, Application US/09381862  
; Patent No. 6245906  
; GENERAL INFORMATION:  
; APPLICANT: Ueyama, Hiroshi  
; APPLICANT: Abe, Kanako  
; APPLICANT: Keshi, Hiroyuki  
; APPLICANT: Matsuhisa, Aki  
; TITLE OF INVENTION: PROBES FOR THE DIAGNOSIS OF INFECTIONS  
; TITLE OF INVENTION: CAUSED BY STREPTOCOCCUS PYOGENES  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive/6300 Sears Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America

```

Query Match      64.08;   Score 16;   DB 4;   Length 6688;
Best Local Similarity 79.23;   Pred. No. 57;
Matches 19;   Conservative 0;   Mismatches 5;   Indels

Oy 1 gacattgcgaagtcacagaatt 24
    ||| |||| |||| |||| |||| ||||
Db 4228 GACGTTGTCGAATTCAGTAAIT 4205

```

```

RESULT 17
US-08-781-562-2/c
; Sequence 2, Application US/08781562
; Patent No. 5763589
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fastseq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,562
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:

```

ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0181 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1188 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: Consensus  
CLONE: Consensus  
US-08-781-562-2

Query Match 63.2%; Score 15.8; DB 1; Length 1188;  
Best Local Similarity 89.5%; Pred. No. 54;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 acattgtcgcaagtccacag 20  
||||||| |||||  
Db 353 ACATTGCGAGAGTCACAG 335

RESULT 18  
US-08-687-379-1  
Sequence 1, Application US/08687379  
Patent No. 5756697  
GENERAL INFORMATION:  
APPLICANT: Hoeger, Thomas  
APPLICANT: Uitsch, Andreas  
APPLICANT: Bach, Alfred  
APPLICANT: Sterrer, Sylvia  
APPLICANT: Lemaire, Hans-Georg  
TITLE OF INVENTION: Subunits of Glutamate Receptors, Their  
TITLE OF INVENTION: Preparation and Their Use  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Keil & Weinkauff  
STREET: 1101 Connecticut Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage  
COMPUTER: IBM AT-compatible, 80286 processor  
OPERATING SYSTEM: MS-DOS version 6.0  
SOFTWARE: WordPerfect version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/687,379  
FILING DATE: 05-AUG-1996  
CLASSIFICATION: 435  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2946 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna to mRNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
DEVELOPMENTAL STAGE: Adult  
TISSUE TYPE: Brain  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 144..2861  
US-08-687-379-1

Query Match 63.2%; Score 15.8; DB 1; Length 2946;  
Best Local Similarity 89.5%; Pred. No. 62;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 3 cattgtcgcaagtccacaga 21  
||||||| |||||  
Db 277 CTTGTGCGCAACTCACAGA 295

RESULT 19  
US-08-687-379-3  
Sequence 3, Application US/08687379  
Patent No. 5756697  
GENERAL INFORMATION:  
APPLICANT: Hoeger, Thomas  
APPLICANT: Uitsch, Andreas  
APPLICANT: Bach, Alfred  
APPLICANT: Sterrer, Sylvia  
APPLICANT: Lemaire, Hans-Georg  
TITLE OF INVENTION: Subunits of Glutamate Receptors, Their  
TITLE OF INVENTION: Preparation and Their Use  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Keil & Weinkauff  
STREET: 1101 Connecticut Avenue  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage  
COMPUTER: IBM AT-compatible, 80286 processor  
OPERATING SYSTEM: MS-DOS version 6.0  
SOFTWARE: WordPerfect version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/687,379  
FILING DATE: 05-AUG-1996  
CLASSIFICATION: 435  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2946 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna to mRNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
DEVELOPMENTAL STAGE: Adult  
TISSUE TYPE: Brain  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 144..2861  
US-08-687-379-3

Query Match 63.2%; Score 15.8; DB 1; Length 2946;  
Best Local Similarity 89.5%; Pred. No. 62;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 3 cattgtcgcaagtccacaga 21  
||||||| |||||  
Db 277 CTTGTGCGCAACTCACAGA 295

RESULT 20  
US-08-254-573-1  
Sequence 1, Application US/08254573  
Patent No. 5610032  
GENERAL INFORMATION:  
APPLICANT: KAMBOJ, Rajender  
APPLICANT: ELIOTT, Candace



STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
PCT-US95-09715-6

Query Match 62.4%; Score 15.6; DB 5; Length 350;  
Best Local Similarity 81.8%; Pred. No. 56;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gacattgtcgaagtccacagaa 22  
||||| ||||| ||  
Db 115 GACATTGGAGCAAGTCACTTAA 94

## RESULT 23

US-08-938-546-1  
Sequence 1, Application US/08938546  
Patent No. 6004556

GENERAL INFORMATION:  
APPLICANT: Fosberry, Andrew P.  
APPLICANT: Lawlor, Elizabeth J.  
APPLICANT: Nicholas, Richard O.  
TITLE OF INVENTION: No. 6004556el rsbu-1  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Dechert Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19103

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/938,546

FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:

FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dickinson, Todd Q  
REGISTRATION NUMBER: 28,354  
REFERENCE/DOCKET NUMBER: P50551  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-994-2252  
TELEFAX: 215-994-2222

TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1002 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-938-546-1

Query Match 62.4%; Score 15.6; DB 3; Length 1002;  
Best Local Similarity 81.8%; Pred. No. 56;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gacattgtcgaagtccacagaa 22  
||||| ||||| ||  
Db 130 GACATTGTCGATATTCACAAA 151

## RESULT 24

US-09-340-812-1  
Sequence 1, Application US/09340812

Patent No. 6111078  
GENERAL INFORMATION:  
APPLICANT: Fosberry, Andrew P.  
APPLICANT: Lawlor, Elizabeth J.  
APPLICANT: Nicholas, Richard O.  
TITLE OF INVENTION: No. 6111078el rsbu-1  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19103

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/340,812

FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/938,546

FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dickinson, Todd Q  
REGISTRATION NUMBER: 28,354  
REFERENCE/DOCKET NUMBER: P50551  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-994-2252  
TELEFAX: 215-994-2222

TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1002 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-340-812-1

Query Match 62.4%; Score 15.6; DB 3; Length 1002;  
Best Local Similarity 81.8%; Pred. No. 66;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gacattgtcgaagtccacagaa 22  
||||| ||||| ||  
Db 130 GACATTGTCGATATTCACAAA 151

## RESULT 25

US-08-938-546-3  
Sequence 3, Application US/08938546  
Patent No. 6004556

GENERAL INFORMATION:  
APPLICANT: Fosberry, Andrew P.  
APPLICANT: Lawlor, Elizabeth J.  
APPLICANT: Nicholas, Richard O.  
TITLE OF INVENTION: No. 6004556el rsbu-1  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Dechert Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19103

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/938,546  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dickinson, Todd Q  
REGISTRATION NUMBER: 28,354  
REFERENCE/DOCKET NUMBER: P50551  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-994-2252  
TELEFAX: 215-994-2222  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1020 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-938-546-3

Query Match 62.4%; Score 15.6; DB 3; Length 1020;  
Best Local Similarity 81.8%; Pred. No. 66;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtccacagaa 22  
||||||| | ||||| |  
Db 148 GACATTGTCGATATTCACAAA 169

RESULT 26  
US-09-340-812-3  
Sequence 3, Application US/09340812  
Patent No. 6111078  
GENERAL INFORMATION:  
APPLICANT: Fosberry, Andrew P.  
APPLICANT: Lawlor, Elizabeth J.  
APPLICANT: Nicholas, Richard O.  
TITLE OF INVENTION: No. 6111078el rsbU-1  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/340,812  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/938,546  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dickinson, Todd Q  
REGISTRATION NUMBER: 28,354  
REFERENCE/DOCKET NUMBER: P50551  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-994-2252  
TELEFAX: 215-994-2222  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:

LENGTH: 1020 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-340-812-3

Query Match 62.4%; Score 15.6; DB 3; Length 1020;  
Best Local Similarity 81.8%; Pred. No. 66;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtccacagaa 22  
||||||| | ||||| |  
Db 148 GACATTGTCGATATTCACAAA 169

RESULT 27  
US-08-938-546-5  
Sequence 5, Application US/08938546  
Patent No. 6004556  
GENERAL INFORMATION:  
APPLICANT: Fosberry, Andrew P.  
APPLICANT: Lawlor, Elizabeth J.  
APPLICANT: Nicholas, Richard O.  
TITLE OF INVENTION: No. 6004556el rsbU-1  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/938,546  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dickinson, Todd Q  
REGISTRATION NUMBER: 28,354  
REFERENCE/DOCKET NUMBER: P50551  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-994-2252  
TELEFAX: 215-994-2222  
TELEX:  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3360 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-938-546-5

Query Match 62.4%; Score 15.6; DB 3; Length 3360;  
Best Local Similarity 81.8%; Pred. No. 80;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtccacagaa 22  
||||||| | ||||| |  
Db 247 GACATTGTCGATATTCACAAA 268

RESULT 28  
US-09-340-812-5



; Sequence 5, Application US/09340812  
; Patent No. 6111078  
; GENERAL INFORMATION:  
; APPLICANT: Fosberry, Andrew P.  
; APPLICANT: Lawlor, Elizabeth J.  
; APPLICANT: Nicholas, Richard O.  
; TITLE OF INVENTION: No. 6111078el rsbu-1  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert Price & Rhoads  
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: US  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/340.812  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/938,546  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dickinson, Todd Q  
; REGISTRATION NUMBER: 28,354  
; REFERENCE/DOCKET NUMBER: P50551  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-994-2252  
; TELEFAX: 215-994-2222  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3360 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-09-340-812-5

Query Match 62.4%; Score 15.6; DB 3; Length 3360;  
Best Local Similarity 81.8%; Pred. No. 80;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtccacagaa 22  
||||||| | ||||| |  
Db 247 GACATTGTCTGATATTCACAAA 268

RESULT 29  
US-09-916-443A-15  
; Sequence 15, Application US/08916443A  
; Patent No. 6001986  
; GENERAL INFORMATION:  
; APPLICANT: Yong Sig KIM  
; APPLICANT: Sun Chung PARK  
; APPLICANT: Soo Kyung OH  
; APPLICANT: Hosuli LEE  
; APPLICANT: Jeong Woo CHO  
; APPLICANT: Chang H. CHUNG  
; TITLE OF INVENTION: Antiviral Proteins, Amarandin 1 and 2, from  
; TITLE OF INVENTION: Amaranthus Viridis, DNAs Encoding Therefrom  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA

; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3+ Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
; SOFTWARE: WordPerfect 6.1 Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/916,443A  
; FILING DATE: 22 AUG 1997  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toffenetti, Judith L.  
; REGISTRATION NUMBER: 39,048  
; REFERENCE/DOCKET NUMBER: 1942/18  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-429-1776  
; TELEFAX: 202-429-0796  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 946 nucleic acids  
; TYPE: nucleotides  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic DNA  
; US-08-916-443A-15  
  
Query Match 61.8%; Score 15.4; DB 3; Length 946;  
Best Local Similarity 76.0%; Pred. No. 82;  
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
  
QY 1 gacattgtcgcaagtccacagaa 25  
||||||| | ||||| | ||||| |  
Db 696 GAAATGTTACAGTCGAGAA 720  
  
RESULT 30  
US-08-916-443A-16  
; Sequence 16, Application US/08916443A  
; Patent No. 6001986  
; GENERAL INFORMATION:  
; APPLICANT: Yong Sig KIM  
; APPLICANT: Sun Chung PARK  
; APPLICANT: Soo Kyung OH  
; APPLICANT: Hosuli LEE  
; APPLICANT: Jeong Woo CHO  
; APPLICANT: Chang H. CHUNG  
; TITLE OF INVENTION: Antiviral Proteins, Amarandin 1 and 2, from  
; TITLE OF INVENTION: Amaranthus Viridis, DNAs Encoding Therefrom  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenyon & Kenyon  
; STREET: 1025 Connecticut Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3+ Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS 6.2  
; SOFTWARE: WordPerfect 6.1 Windows  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/916,443A  
; FILING DATE: 22 AUG 1997  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toffenetti, Judith L.  
; REGISTRATION NUMBER: 39,048  
; REFERENCE/DOCKET NUMBER: 1942/18  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-429-1776  
; TELEFAX: 202-429-0796



APPLICANT: Tan, Jimmy  
APPLICANT: Zavodny, Paul  
TITLE OF INVENTION: Mammalian TNF-a Convertases  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Schering-Plough Corporation  
STREET: 2000 Galloping Hill Road  
CITY: Kenilworth  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07033-0530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 7.1  
SOFTWARE: Microsoft Word 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/889,909A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/021,710  
FILING DATE: 12-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Dulak, No. 5853977man C.  
REGISTRATION NUMBER: 31,608  
REFERENCE/DOCKET NUMBER: JB0601  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 908 298 2906  
TELEFAX: 908 298 5388  
TELEX:  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2352 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-889-909A-21

Query Match 61.6%; Score 15.4; DB 2; Length 2352;  
Best Local Similarity 76.0%; Pred. No. 94;  
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gacattgtcgaagtccacagaatta 25  
Db 2351 GACTTATCGAGCTCTCAGATT 2327

RESULT 34  
US-08-665-259-24  
Sequence 24, Application US/08665259  
Patent No. 6028173  
GENERAL INFORMATION:  
APPLICANT: Landes, Gregory M.  
APPLICANT: Burn, Timothy C.  
APPLICANT: Connors, Timothy D.  
APPLICANT: Dackowski, William R.  
APPLICANT: Van Raay, Terence J.  
APPLICANT: Klingner, Katherine W.  
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,  
METHODS OF MAKING AND USING SAME  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENZYME CORPORATION  
STREET: One Mountain Road  
CITY: Framingham  
STATE: Massachusetts  
COUNTRY: United States of America  
ZIP: 01701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/665,259  
FILING DATE: 17-JUN-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Dugan, Deborah A.  
REGISTRATION NUMBER: 37,315  
REFERENCE/DOCKET NUMBER: IGS-9.1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (508) 872-8400  
TELEFAX: (508) 872-5415  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5894 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2..5053  
US-08-665-259-24

Query Match 61.6%; Score 15.4; DB 3; Length 5894;  
Best Local Similarity 76.0%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gacattgtcgaagtccacagaatta 25  
Db 4259 GCCTTTGTGGGGGTACAGATCA 4283

RESULT 35  
US-08-762-500-24  
Sequence 24, Application US/08762500  
Patent No. 6030806  
GENERAL INFORMATION:  
APPLICANT: Landes, Gregory M.  
APPLICANT: Burn, Timothy C.  
APPLICANT: Connors, Timothy D.  
APPLICANT: Dackowski, William R.  
APPLICANT: Van Raay, Terence J.  
APPLICANT: Klingner, Katherine W.  
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,  
METHODS OF MAKING AND USING SAME  
NUMBER OF SEQUENCES: 83  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENZYME CORPORATION  
STREET: One Mountain Road  
CITY: Framingham  
STATE: Massachusetts  
COUNTRY: United States of America  
ZIP: 01701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/762,500  
FILING DATE: 09-DEC-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/665,259  
FILING DATE: 17-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/10469  
FILING DATE: 17-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Dugan, Deborah A.

```
; LENGTH: 6525 base pairs
```

Db 414 CATGTCGTTAGTTTCAGAT

Qy 3 cattgtcgcaagtcacagaatta 25  
||||| ||| |||||  
Db 414 CATTGTCGTTAGTTTCAGATTTA 392

```
RESULT 38
US-08-821-324-7/C
; Sequence 7, Application US/08821324
; Patent No. 6231869
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND THERAPY
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,324
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.439
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 530 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Ehrlichia
;
; US-08-821-324-7
;
; Query Match 60.0%; Score 15; DB 4; Length 530;
; Best Local Similarity 78.3%; Pred. No. 1.2e+02;
; Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
;
; QY 3 cattgtcgcaagtcacagaatta 25
; ||||| ||| ||||| |||
; DB 414 CATTGCGTTAGTTTCAGATTTA 392
;
; RESULT 39
US-08-116-098-1
; Sequence 1, Application US/08116098
; Patent No. 5428131
; GENERAL INFORMATION:
; APPLICANT: Trent, Jonathan D.
; APPLICANT: Horwich, Arthur L.
; TITLE OF INVENTION: Archaeobacterial Chaperonin-Mediated
; TITLE OF INVENTION: Protein Stabilization
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/116,098
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/756627
FILING DATE: 09-SEP-1991
APPLICATION NUMBER: US 07/721974
FILING DATE: 27-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/673158
FILING DATE: 18-MAR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/261573
FILING DATE: 24-OCT-1988
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: YU102
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-815-6508
TELEFAX: 404-815-6555
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1749 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Sulfolobus shibatae
FEATURE:
NAME/KEY: promoter
LOCATION: 1..93
;
; US-08-116-098-1
;
; Query Match 60.0%; Score 15; DB 1; Length 1749;
; Best Local Similarity 78.3%; Pred. No. 1.4e+02;
; Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
;
; QY 3 cattgtcgcaagtcacagaatta 25
; ||||| ||||| ||||| |||||
; DB 653 CAGTGGCACAAAGTAGCAGAATTA 675
;
; RESULT 40
US-08-375-709-14
; Sequence 14, Application US/08375709
; Patent No. 5683898
; GENERAL INFORMATION:
; APPLICANT: YAZAWA, Kazunaga
; APPLICANT: YAMADA, Akiko
; APPLICANT: KATO, Seishi
; APPLICANT: KONDO, Kiyosi
; TITLE OF INVENTION: Gene Coding For Eicosapentaenoic Acid
; TITLE OF INVENTION: Synthesizing Enzymes and Process for Production of
; TITLE OF INVENTION: Eicosapentaenoic Acid
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/375,709
; FILING DATE: 20-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/178,251
; FILING DATE: 14-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 4-147945
; FILING DATE: 15-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: WEGNER, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 53466/150/AAOK
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6012 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Shewanella putrefaciens SCRC-2874 (FERM
; ORGANISM: BP-1625)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..6012
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..6012
; US-08-375-709-14

```

Query Match 60.0%; Score 15; DB 1; Length 6012;

Best Local Similarity 78.3%; Pred. No. 1.7e+02;

Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

QY 1 gacattgtcgcaagtccacagaat 23
|| |||| |||| |||| ||||
Db 2868 GAAATTGGCGCAAGCGACGACAT 2890

```

#### RESULT 41

US-08-752-929-14

; Sequence 14, Application US/08752929

; Patent No. 5798259

; GENERAL INFORMATION:

; APPLICANT: YAZAWA, Kazunaga

; APPLICANT: YAMADA, Akiko

; APPLICANT: KATO, Seishi

; APPLICANT: KONDO, Kiyosi

; TITLE OF INVENTION: Gene Coding for Eicosapentaenoic Acid Synthesizing

; TITLE OF INVENTION: Enzymes and Process for Production of Eicosapentaenoic

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752,929
; FILING DATE: 20-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/375,709
; FILING DATE: 20-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/178,251
; FILING DATE: 14-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 4-147945
; FILING DATE: 15-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: WEGNER, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 53466/150/AAOK
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6012 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Shewanella putrefaciens SCRC-2874 (FERM
; ORGANISM: BP-1625)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..6012
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..6012
; US-08-752-929-14

```

Query Match

Best Local Similarity 60.0%; Score 15; DB 1; Length 6012;

Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

QY 1 gacattgtcgcaagtccacagaat 23
|| |||| |||| |||| ||||
Db 2868 GAAATTGGCGCAAGCGACGACAT 2890

```

#### RESULT 42

PCI-US95-11859-2/c

; Sequence 2, Application PC/TUS9511859

; GENERAL INFORMATION:

; APPLICANT: JANSEN, KATHRYN U.

; APPLICANT: HOFMANN, KATHRYN J.

; TITLE OF INVENTION: DNA ENCODING HUMAN PAPILLOMA VIRUS TYPE

; TITLE OF INVENTION: 6A

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CHRISTINE E. CARTY

; STREET: 126 E. LINCOLN AVENUE - P.O. BOX 2000

; CITY: RAHWAY

; STATE: NJ

; COUNTRY: USA

; ZIP: 07065-0907

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/11859  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/310,468  
FILING DATE: 22-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: CARTY, CHRISTINE E.  
REGISTRATION NUMBER: 36,099  
REFERENCE/DOCKET NUMBER: 19307 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-6734  
TELEFAX: (908) 594-4720  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8010 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
PCT-US95-11859-2

Query Match 60.0%; Score 15; DB 5; Length 8010;  
Best Local Similarity 78.3%; Pred. No. 1.8e+02;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 cattgtcgcaagtcacagaatta 25  
||||| ||||| ||||| |||||  
Db 211 CATCTTGCACAAACACACAAATTA 189

## RESULT 43

US-08-375-709-1  
Sequence 1, Application US/08375709  
Patent No. 5683898  
GENERAL INFORMATION:  
APPLICANT: YAZAWA, Kazunaga  
APPLICANT: YAMADA, Akiko  
APPLICANT: KATO, Seishi  
APPLICANT: KONDO, Kiyosi  
TITLE OF INVENTION: Gene Coding For Eicosapentaenoic Acid  
TITLE OF INVENTION: Synthesizing Enzymes and Process for Production of  
TITLE OF INVENTION: Eicosapentaenoic Acid  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/375,709  
FILING DATE: 20-JAN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/178,251  
FILING DATE: 14-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-147945  
FILING DATE: 15-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: WEGNER, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 53466/150/AAOK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 37895 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Shewanella putrefaciens SCRC-2874 (FERM  
ORGANISM: BP-1625)  
US-08-375-709-1

Query Match 60.0%; Score 15; DB 1; Length 37895;  
Best Local Similarity 78.3%; Pred. No. 2.2e+02;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaat 23  
||||| ||||| ||||| |||||  
Db 27385 GAAATGGCGCAAGCGACAGCAT 27407

## RESULT 44

US-08-752-929-1  
Sequence 1, Application US/08752929  
Patent No. 5798259  
GENERAL INFORMATION:  
APPLICANT: YAZAWA, Kazunaga  
APPLICANT: YAMADA, Akiko  
APPLICANT: KATO, Seishi  
APPLICANT: KONDO, Kiyosi  
TITLE OF INVENTION: Gene Coding for Eicosapentaenoic Acid Synthesizing  
TITLE OF INVENTION: Enzymes and Process for Production of Eicosapentaenoic  
TITLE OF INVENTION: Acid  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/752,929  
FILING DATE: 20-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/375,709  
FILING DATE: 20-JAN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/178,251  
FILING DATE: 14-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-147945  
FILING DATE: 15-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: WEGNER, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 53466/150/AAOK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 37895 base pairs

; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: Shewanella putrefaciens SCRC-2874 (FERM  
; ORGANISM: BP-1625)  
US-08-752-929-1

Query Match 60.0%; Score 15; DB 1; Length 37895;  
Best Local Similarity 78.3%; Pred. No. 2.2e+02;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaat 23  
|| |||| ||||| |||||  
Db 27385 GAAATTGGCGCAAGCGACGACAT 27407

## RESULT 45

US-09-090-793-1  
; Sequence 1, Application US/09090793  
; Patent No. 6140486  
; GENERAL INFORMATION:  
; APPLICANT: Calgene, LLC  
; TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression  
; FILE REFERENCE: CGNE.131.01US  
; CURRENT APPLICATION NUMBER: US/09/090,793  
; EARLIER FILING DATE: 1998-06-04  
; EARLIER APPLICATION NUMBER: 60/048,650  
; EARLIER FILING DATE: 1997-06-04  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 37895  
; TYPE: DNA  
; ORGANISM: Shewanella putrefaciens  
US-09-090-793-1

Query Match 60.0%; Score 15; DB 3; Length 37895;  
Best Local Similarity 78.3%; Pred. No. 2.2e+02;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaat 23  
|| |||| ||||| |||||  
Db 27385 gaaattggcgcaagcgacagcat 27407

Search completed: October 9, 2001, 11:39:28  
Job time: 1883 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:08:05 ; Search time 5930.9 Seconds  
(without alignments)  
39.846 Million cell updates/sec

Title: US-09-396-196F-1  
Perfect score: 25  
Sequence: 1 gacatgtcgcaagtcacagaatta 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est33:\*  
26: gb\_est34:\*  
27: gb\_est35:\*  
28: gb\_est36:\*  
29: gb\_est37:\*  
30: gb\_est38:\*  
31: gb\_est39:\*  
32: gb\_est40:\*  
33: em\_estba:\*  
34: em\_estfun:\*  
35: em\_esthum1:\*  
36: em\_esthum2:\*  
37: em\_esthum3:\*  
38: em\_esthum4:\*  
39: em\_esthum5:\*  
40: em\_esthum6:\*  
41: em\_esthum7:\*  
42: em\_esthum8:\*  
43: em\_esthum9:\*

44: em\_esthum10:\*  
45: em\_esthum11:\*  
46: em\_esthum12:\*  
47: em\_esthum13:\*  
48: em\_esthum14:\*  
49: em\_esthum15:\*  
50: em\_esthum16:\*  
51: em\_esthum17:\*  
52: em\_esthum18:\*  
53: em\_esthum19:\*  
54: em\_esthum20:\*  
55: em\_esthum21:\*  
56: em\_esthum22:\*  
57: em\_esthum23:\*  
58: em\_esthum24:\*  
59: em\_esthum25:\*  
60: em\_esthum26:\*  
61: em\_esthum27:\*  
62: em\_esthum28:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estin5:\*  
68: em\_estom1:\*  
69: em\_estom2:\*  
70: em\_estov1:\*  
71: em\_estov2:\*  
72: em\_estpl1:\*  
73: em\_estpl2:\*  
74: em\_estpl3:\*  
75: em\_estpl4:\*  
76: em\_estpl5:\*  
77: em\_estpl6:\*  
78: em\_estpl7:\*  
79: em\_estpl8:\*  
80: em\_estpl9:\*  
81: em\_estpl10:\*  
82: em\_estro1:\*  
83: em\_estro2:\*  
84: em\_estro3:\*  
85: em\_estro4:\*  
86: em\_estro5:\*  
87: em\_estro6:\*  
88: em\_estro7:\*  
89: em\_estro8:\*  
90: em\_estro9:\*  
91: em\_estro10:\*  
92: em\_estro11:\*  
93: em\_estro12:\*  
94: em\_estro13:\*  
95: em\_estro14:\*  
96: em\_estro15:\*  
97: em\_estro16:\*  
98: em\_estro17:\*  
99: em\_estro18:\*  
100: em\_estro19:\*  
101: em\_estro20:\*  
102: gb\_est25:\*  
103: gb\_est26:\*  
104: gb\_est27:\*  
105: gb\_est28:\*  
106: gb\_est29:\*  
107: gb\_est30:\*  
108: gb\_est31:\*  
109: gb\_est32:\*  
110: gb\_est41:\*  
111: gb\_est42:\*  
112: gb\_est43:\*  
113: gb\_est44:\*  
114: gb\_est45:\*  
115: gb\_est46:\*  
116: gb\_est47:\*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	19.2	76.8	576	233	AQ771996	HS_5413_B
C 2	18.6	74.4	494	227	AQ335616	HS_5015_B
C 3	18.6	74.4	596	238	AZ083812	RPCI-23-2
C 4	18.6	74.4	869	112	AW155118	mgie0002K
C 5	18.2	72.8	169	190	W18068	mb85d03.r1
C 6	18.2	72.8	302	145	BF172143	PC14225_M
C 7	18.2	72.8	357	105	AL042323	DKFZp4340
C 8	18.2	72.8	357	140	BE800706	sg97c06.y
C 9	18.2	72.8	387	256	B48721	RPCI11-3B21
C 10	18.2	72.8	388	120	AW781150	sl89g06.y
C 11	18.2	72.8	444	189	T83852	yd66b07.s1
C 12	18.2	72.8	454	159	N58785	yv76c09.s1
C 13	18.2	72.8	506	15	AI017333	ov31h07.x
C 14	18.2	72.8	557	228	AQ437316	HS_5083_A
C 15	18.2	72.8	567	240	AZ225990	RPCI-23-5
C 16	18.2	72.8	709	223	AQ020375	CIT-HSP-2
C 17	18.2	72.8	751	138	BE658717	GM700007A
C 18	17.8	71.2	352	225	AQ238325	RPCI11-69
C 19	17.8	71.2	379	187	R21156	yh20a04.r1
C 20	17.8	71.2	382	163	BE101686	UI-R-R01-
C 21	17.6	70.4	194	121	AW833543	QV4-TT000
C 22	17.6	70.4	203	121	AW833554	QV4-TT000
C 23	17.6	70.4	225	111	AW057146	ca01d05.y
C 24	17.6	70.4	232	121	AW833631	QV4-TT000
C 25	17.6	70.4	232	121	AW833768	QV4-TT000
C 26	17.6	70.4	274	121	AW833615	QV4-TT000
C 27	17.6	70.4	403	227	AQ385232	RPCI11-14
C 28	17.6	70.4	431	226	AQ344476	RPCI11-10
C 29	17.6	70.4	460	105	AL040883	AL040883
C 30	17.6	70.4	499	224	AQ142135	HS_3157_B
C 31	17.6	70.4	517	225	AQ215052	HS_3244_A
C 32	17.6	70.4	527	241	AZ337152	IM0067B09
C 33	17.6	70.4	599	244	AZ476585	IM0295H03
C 34	17.6	70.4	609	241	AZ325063	IM0047J15
C 35	17.6	70.4	615	230	AQ062696	HS_2116_A
C 36	17.6	70.4	643	226	AQ306368	HS_2050_A
C 37	17.6	70.4	659	110	AW038199	EST279856
C 38	17.6	70.4	661	228	AQ389869	RPCI11-14
C 39	17.6	70.4	667	241	AZ289880	RPCI-23-1
C 40	17.6	70.4	719	241	AZ334568	IM0084F02
C 41	17.4	69.6	531	233	AQ060620	HS_3237_A
C 42	17.4	69.6	612	228	AQ403329	HS_5061_B
C 43	17.4	69.6	622	242	AZ385697	IM0144M15
C 44	17.2	68.8	163	108	AV006577	AV006577
C 45	17.2	68.8	317	244	AZ456884	IM0260B02

## ALIGNMENTS

RESULT	1				
AQ771996/c					
LOCUS		576 bp	DNA	GSS	29-JUL-1999
DEFINITION	HS_5413_B2_F07_SPEE RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate-989 Col-14 Row-L, DNA sequence.				
ACCESSION	AQ771996				
VERSION	AQ771996.1	GI:5651724			
KEYWORDS	GSS.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 576) Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.				
TITLE	Sequence-tagged connectors: A sequence approach to mapping and				

## JOURNAL MEDLINE COMMENT

scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.bufo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.bufo.edu/ordering\_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu  
Plate: 989 row: L column: 14  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 576.

## FEATURES source

Location/Qualifiers  
1..576  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate=989 Col=14 Row=L"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"  
BASE COUNT 167 a 126 c 109 g 168 t 6 others  
ORIGIN

Query Match 76.8%; Score 19.2; DB 233; Length 576;  
Best Local Similarity 87.5%; Pred. No. 66;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 acattgtcgaagtacagaatta 25  
||||| ||||| ||||| |||||

Db 219 ACATTGTCTCAAGCCACACAATTA 196

## RESULT 2

AQ335616/c 494 bp DNA GSS 06-MAR-1999  
LOCUS HS\_5015\_B2\_C08\_SP6E RPCI11 Human Male BAC Library Homo sapiens  
DEFINITION genomic clone Plate-591 Col-16 Row-F, DNA sequence.  
ACCESSION AQ335616  
VERSION AQ335616.1 GI:4142595  
KEYWORDS GSS.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE AUTHORS

1 (bases 1 to 494)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

## TITLE

Sequence-tagged connectors: A sequence approach to mapping and

## JOURNAL MEDLINE COMMENT

scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com).

BAC end Web Server: <http://www.htsc.washington.edu>  
 Plate: 591 row: F column: 16

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 494.

Location/Qualifiers

1. .494

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="plate=591 Col=16 Row=F"

/clone\_lib="RPCI11 Human Male BAC Library"

/sex="Male"

/cell\_type="Lymphocytes"

/note="vector: pBACE3.6; RPCI11 Human Male BAC Library"

207 a 82 c 91 g 112 t 2 others

#### BASE COUNT

#### ORIGIN

Query Match 74.4%; Score 18.6; DB 227; Length 494;  
 Best Local Similarity 84.0%; Pred. No. 1.3e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaatta 25

Db 373 GACATTGTGACGATGCACAGATCTA 349

#### RESULT 3

#### LOCUS

DEFINITION AZ083812 596 bp DNA GSS 08-MAY-2000

RPCI-23-22N4.TJ RPCI-23 Mus musculus genomic clone RPCI-23-22N4,

DNA sequence.

#### ACCESSION

#### VERSION

AZ083812 1 GI:7725545

#### KEYWORDS

#### SOURCE

#### ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 596)

Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret

,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.

and Fraser,C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: [szhao@tigr.org](mailto:szhao@tigr.org)

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

([pieter@dejong.med.buffalo.edu](mailto:pieter@dejong.med.buffalo.edu)). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)

or from Resea ch Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC end page:

[http://www.tigr.org/db/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/db/bac_ends/mouse/bac_end_intro.html)

Plate: 22 row: N column: 4

Seq primer: SP6

Class: BAC ends.

#### FEATURES

#### source

1. .596

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-23-22N4"

/clone\_lib="RPCI-23"

/sex="Female"

/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site\_1:

EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested

with a combination of EcoRI and EcoRI Methylase. Size  
 selected DNA was cloned into the pBACE3.6 vector at the  
 EcoRI sites. The ligation products were transformed into  
 DH10B electrocompetent cells (BRL Life Technologies).

BASE COUNT 174 a 108 c 95 g 219 t

#### ORIGIN

Query Match 74.4%; Score 18.6; DB 238; Length 596;

Best Local Similarity 84.0%; Pred. No. 1.3e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaatta 25

Db 561 GGCATTGTCCAAAGTCACAGAACTA 585

#### RESULT 4

#### LOCUS

DEFINITION

AW155118 869 bp mRNA EST 03-NOV-1999

mgie0002K13f Rice blast infection stage cDNA library Oryza

sativa/Pyricularia oryzae mixed EST library cDNA clone mgie0002K13f

, mRNA sequence.

ACCESSION AW155118

VERSION AW155118.1 GI:6223987

KEYWORDS EST.

SOURCE Oryza sativa/Pyricularia oryzae mixed EST library.

ORGANISM Oryza sativa/Pyricularia oryzae mixed EST library

Eukaryota; mixed EST libraries.

REFERENCE 1 (bases 1 to 869)

AUTHORS Rauyaree,R., Choi,W. and Dean,R.A.

TITLE Identification and characterization of genes expressed by the rice

blast pathogen and rice during infection stage

Unpublished (1999)

JOURNAL

COMMENT

Contact: Dean,R.A.

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson University, Clemson, SC 29634, USA

Tel: 864 656 5737

Fax: 864 656 4293

Email: [rdean@clemson.edu](mailto:rdean@clemson.edu)

Seq primer: T3 primer (AATTAACCTCACTAAAGGG)

High quality sequence stop: 257.

FEATURES

source

1. .869

/organism="Oryza sativa/Pyricularia oryzae mixed EST

library"

/strain="70-15"

/db\_xref="taxon:105664"

/clone="mgie0002K13f"

/clone\_lib="Rice blast infection stage cDNA library"

/dev\_stage="Infection stage at 48 hour post-inoculation"

/note="Vector: pBlueScriptII SK(+) Vector; Rice

blast-infected leaves at 48 hour post-inoculation mRNA for

cDNA library construction."

BASE COUNT 236 a 179 c 173 g 271 t

ORIGIN

Query Match 74.4%; Score 18.6; DB 112; Length 869;

Best Local Similarity 84.0%; Pred. No. 1.4e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaatta 25

Db 520 GACATGGACCGCTAGTCACATAATTA 544

#### RESULT 5

#### LOCUS

DEFINITION

W18068 169 bp mRNA EST 10-SEP-1996

mb85d03.r1 Soares mouse p3NNF19.5 Mus musculus cDNA clone

IMAGE:336197 5' similar to PIR:A48080 A48080 basic helix-loop-helix

1 (bases 1 to 302)  
 Claudio J.O., Tang H., Khan E.M., Voralia M., Li Z., Cukerman E.,  
 Francisco-Palau O., Liew C.C. and Stewart A.K.  
 The transcriptional phenotype of myeloma cells  
 Unpublished (2000)  
 Contact: A. Keith Stewart, M.D.  
 Oncology Research  
 University Health Network  
 610 University Ave., 5-126, Toronto, Ontario, M5G 2M9, Canada  
 Tel: (416) 946-4639  
 Fax: (416) 946-6546  
 Email: k.stewart@utoronto.ca

PCR Primers  
 FORWARD: 5'-GCCAAGCTCAAAATTAACCTCACTAAAGG-3'  
 BACKWARD: 5'-CCAGTGAATGTAACTAGCACTACTATAGGCG-3'  
 Seq primer: 5'-GAAATTAACCTCACTAAAG-3'

Location/Qualifiers  
 1..302  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Myeloma (PCL) cDNA library"  
 /sex="male"  
 /tissue\_type="Blood"  
 /cell\_type="myeloma"  
 /dev\_stage="Plasma cell leukemia"  
 /note="vector: Lambda Zap Express: Site 1: EcoRI; Site 2:  
 XhoI; mRNA was purified from plasma cell leukemia  
 patient's peripheral blood containing >95% myeloma. An  
 oligo d(T)18 primer containing XhoI restriction site was  
 used to prime first strand synthesis using M-MLV reverse  
 transcriptase. To protect the cDNAs from XhoI digestion  
 subsequent cloning step, the nucleotide analogue  
 5-methyl-dCTP was added to the nucleotide mixture and  
 la-32P dATP was added to monitor the quantity and quality  
 of first strand synthesis. After second-strand synthesis  
 and blunting of cDNA termini, EcoRI adapters were ligated  
 , followed by kinase treatment and digestion with XhoI.  
 The cDNAs were then size-fractionated using Sephacryl  
 S-500 column and then ligated into EcoRI and XhoI digested  
 Lambda Zap Express vector. The ligation product was  
 packaged using Gigapack II packaging extract. The library  
 had primary titre of approx. 1x10<sup>6</sup>. Clones from the  
 primary library were randomly selected for single pass  
 sequencing."

108 a 53 c 64 g 77 t

Query Match 72.8% Score 18.2; DB 145; Length 302;  
 Best Local Similarity 87.08; Pred. No. 1.8e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0

2 acattgtcgcaagtcacagaatt 24  
 ||||| ||||| |||||  
 Db 209 ACATTGTAGCAAGTCTCAGAATT 187

RESULT 7  
 AL042323/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 357)  
 Ottenwaelder,B., Obermaier,B., Mewes,H.W., Gassenhuber,J. and  
 Wiemann,S.  
 EST (Ottenwaelder, et al.)

JOURNAL  
COMMENT

Unpublished (1999)  
On Jul 8, 1999 this sequence version replaced gi:5421665.  
Contact: Ottenwaelder B  
MIPS

Am Klopferspitz 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;  
sequenced by MedGenomix (Martinsried/Germany) within the cDNA  
sequencing consortium of the German Genome Project. No sl sequence  
available.

This clone (DKFZp434O1120) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

## FEATURES

source

```
1..357
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp434O1120"
/clone_lib="434 (synonym: htes3)"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
```

BASE COUNT 132 a 74 c 60 g 91 t

91 t

Query Match 72.8%; Score 18.2; DB 105; Length 357;  
Best Local Similarity 87.0%; Pred. No. 1.9e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 acattgtcgcaagtcacagaatt 24

```
|||||
|||||
|||||
|||||
```

Db 206 ACTTTTCCCAAGTCACAGAATT 184

## RESULT 8

BE800706

LOCUS

sq97c06.y1 Gm-c1049 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
Gm-c1049-1091 5' similar to TR:082150 082150 CHLOROPLAST FISH  
PROTEASE. ;, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Glycine max  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.

REFERENCE

AUTHORS

Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna  
,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,  
Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers  
,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk  
,R., Ritter,E., Kohn,S., Shin,T., Jackson,K., Cardenas,M., McCann  
,R., Waterston,R. and Wilson,R.  
Public Soybean EST Project  
Unpublished (1999)

TITLE

JOURNAL

COMMENT

Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: est@watson.wustl.edu  
This clone is available through: Genome Systems, Inc. 4633 World  
Parkway Circle St. Louis, Missouri 63134 For further information  
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
427-3324 or contact: clones@genomesystems.com or  
info@genomesystems.com web site: www.genomesystems.com  
Putative full length read

## FEATURES

source

```
1..357
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1049-1091"
/clone_lib="Gm-c1049"
/tissue_type="whole seedlings of greenhouse grown plants"
/dev_stage="3 week old"
/lab_host="DH10B"
```

/note="Vector: pBluescript II SK+; Site\_1: EcoRI; Site\_2:  
XhoI; The Clark NIL was constructed and seed was provided  
by Dr. J. Specht, University of Nebraska (Shoemaker and  
Specht, 1995). The cDNA library was constructed from mRNA  
isolated from whole seedlings of 3 week old greenhouse  
grown plants. Complementary DNA was synthesized from mRNA  
using a primer consisting of a poly(dT) sequence with a  
XhoI restriction site and a 3' anchor. EcoRI adapters were  
ligated to the blunt-ended cDNA fragments followed by XhoI  
digestion. The cDNA fragments were directionally cloned  
into the EcoRI-XhoI restriction site of the pBluescript  
vector. The ligated cDNA fragments were transformed into  
DH10B host cells (GibcoBRL). The library was constructed  
in cooperation with Dr. Paul Keim's laboratory at Northern  
Arizona University."

BASE COUNT 169 a 53 c 47 g 88 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 140; Length 357;

Best Local Similarity 87.0%; Pred. No. 1.9e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 cattgtcgcaagtcacagaatta 25

```
|||||
|||||
|||||
|||||
```

Db 132 CATAACGGCAAGTCACAGAATTA 154

## RESULT 9

B48721/c

LOCUS

B48721 387 bp DNA GSS 08-APR-1999  
RPC111-3B21.TP RPCI-11 Homo sapiens genomic clone RPCI-11-3B21, DNA  
sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 387)  
Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden  
,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter  
,J.C.

Use of BAC End Sequences for Sequence-Ready Map Building

TITLE

JOURNAL

COMMENT

Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from  
Research Genetics (<http://inforesgen.com>). BAC end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: SP6  
Class: BAC ends.

FEATURES Location/Qualifiers

```

source
1. .387
/organism="Homo sapiens"
/db_xref="GDB:7500812"
/db_xref="taxon:9606"
/clone_lib="RPCI-11-3B21"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"
BASE COUNT      130 a      60 c      91 g      106 t
ORIGIN

Query Match      72.8%; Score 18.2; DB 256; Length 387;
Best Local Similarity 87.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 acattgcgcagtcacagaatt 24
||||| ||||| |||||
Db 160 ACATTGTAGCAAGTGTACAGAAIT 138

RESULT 10
AW781150
LOCUS      AW781150      388 bp      mRNA      EST      12-MAY-2000
DEFINITION      sl89g06.y1 Gm-cl037 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl037-1235 5', similar to TR:082150 O82150 CHLOROPLAST FTSH
PROTEASE. ;, mRNA sequence.
ACCESSION      AW781150
VERSION        AW781150.1 GI:7795753
KEYWORDS
SOURCE        soybean.
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eucosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE      1 (bases 1 to 388)
AUTHORS      Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
,R., Waterston,R. and Wilson,R.
TITLE        Public Soybean EST Project
JOURNAL
COMMENT      Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
High quality sequence stop: 375.
Location/Qualifiers
1. .388
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl037-1235"
/clone_lib="Gm-cl037"
/tissue_type="fully expanded leaves of greenhouse grown
plants"
/dev_stage="2 week old"
/lab_host="DH10B"
/note="Vector: pSPOR1; Site_1: NotI; Site_2: SalI; This
cDNA library was constructed from mRNA isolated from fully
expanded leaves of greenhouse grown plants that were 2
weeks old. The library was prepared using the Life

```

```

Technologies pSuperscript cDNA library construction kit.
Complementary DNA was synthesized from mRNA using a
poly(dT) sequence with a NotI restriction site. SalI
linkers adapters were ligated to the blunt-ended cDNA
fragments followed by NotI digestion. The cDNA fragments
were directionally cloned into the NotI-SalI restriction
site of the pSPORT1 vector. The ligated cDNA fragments
were transformed into E.coli Electro-Max DH10B host
cells. This library was constructed in the laboratory of
Dr. Lila Vodkin by Anu Khanna at the University of
Illinois at Urbana- Champaign. email: l-vodkin@uiuc.edu"
BASE COUNT      123 a      67 c      92 g      106 t
ORIGIN

Query Match      72.8%; Score 18.2; DB 120; Length 388;
Best Local Similarity 87.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 cattgtcgcagtcacagaatt 25
||||| ||||| |||||
Db 328 CATACACCAAGTCACAGAAITTA 350

RESULT 11
T83852
LOCUS      T83852      444 bp      mRNA      EST      16-MAR-1995
DEFINITION      yd66b07.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:113173 3' similar to gb:IM87934:HUMALUD43 Human carcinoma
cell-derived Alu RNA transcript, (cRNA); gb:X54150_rnal
IMMUNOGLOBULIN ALPHA FC RECEPTOR PRECURSOR (HUMAN); contains Alu
repetitive element; contains 11 repetitive element ;, mRNA sequence.
ACCESSION      T83852
VERSION        T83852.1 GI:712140
KEYWORDS      human.
SOURCE        Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 444)
AUTHORS      Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE        The WashU-Merck EST Project
JOURNAL
COMMENT      Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 939
High quality sequence stops: 334 Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 939 Std Error: 0.00
Seq primer: T3
High quality sequence stop: 334.
Location/Qualifiers
1. .444
/organism="Homo sapiens"
/db_xref="GDB:468790"
/db_xref="taxon:9606"
/clone="IMAGE:113173"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: p7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACGCGAAGAAATTAATAAGATCTTTTTTTTTTTTTTTT 3'],

```

double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 108 a 69 c 100 g 162 t 5 others  
ORIGIN

Query Match 72.8%; Score 18.2; DB 189; Length 444;  
Best Local Similarity 87.0%; Pred. No. 1.9e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 acattgtcgcaagtcacagaatt 24  
||||| ||||| ||||| |||||  
Db 188 ATATTGTTGCAAGTAACAGAATT 210

## RESULT 12

N58785 454 bp mRNA EST 30-JAN-1997  
LOCUS YV76C09.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone  
IMAGE:248656 3' similar to contains Alu repetitive element; contains  
element L1 repetitive element ;, mRNA sequence.

ACCESSION N58785  
VERSION N58785.1 GI:1202675  
KEYWORDS EST.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 454)

## REFERENCE

AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins,  
M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore,  
B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,  
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,  
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)

## TITLE

JOURNAL Washington University School of Medicine  
MEDLINE 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
COMMENT Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu

Contact: Wilton RK

Washington University School of Medicine

444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: estewatson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 916 Std Error: 0.00

Seq primer: mob.REGA+ET

High quality sequence stop: 1.

## FEATURES

## source

1. .454  
/organism="Homo sapiens"  
/db\_xref="GDB:3797902"  
/db\_xref="taxon:9606"  
/clone="IMAGE:248656"  
/clone\_lib="Soares fetal liver spleen INFLS"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"

/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)  
with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
1st strand cDNA was primed with a Pac I - oligo(dT) primer  
[5' ACTGGAAGAATAATTAAGATCTTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M. Fatima Bonaldo."

116 a 65 c 104 g 168 t 1 others

## BASE COUNT

## ORIGIN

Query Match 72.8%; Score 18.2; DB 159; Length 454;  
Best Local Similarity 87.0%; Pred. No. 1.9e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 acattgtcgcaagtcacagaatt 24  
||||| ||||| ||||| |||||  
Db 191 ATATTGTTGCAAGTAACAGAATT 213

## RESULT 13

AI017333 506 bp mRNA EST 27-AUG-1998  
LOCUS OV31h07.x1 Soares\_testis\_NHT Homo sapiens cDNA clone IMAGE:1638973  
DEFINITION 3', mRNA sequence.

ACCESSION AI017333  
VERSION AI017333.1 GI:3231669  
KEYWORDS EST.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 506)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo,  
Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Insert Length: 1675 Std Error: 0.00

Seq primer: -40ml3 fwd. Et from Amersham

High quality sequence stop: 451.

## FEATURES

## source

1. .506  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1638973"  
/clone\_lib="Soares\_testis\_NHT"  
/sex="male"  
/lab\_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA

was prepared from mRNA obtained from Clontech Laboratories

, Inc., and primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGAGCGGCCCAATTTTTTTTTTTT 3'].

Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT7T3 vector. Library

went through one round of normalization to Cot5, and was

constructed by Bento Soares and M. Fatima Bonaldo."

156 a 84 c 95 g 170 t 1 others

## BASE COUNT

## ORIGIN

Query Match 72.8%; Score 18.2; DB 15; Length 506;  
Best Local Similarity 87.0%; Pred. No. 2e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 acattgtcgcaagtcacagaatt 24  
||||| ||||| ||||| |||||  
Db 135 ACTTTTCCCAAGTCACAGAATT 157

## RESULT 14

## A0437316



LOCUS	AQ437316	557 bp	DNA	GSS	31-MAR-1999
DEFINITION	HS_5083_A2_B06_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=659 Col=12 Row=C, DNA sequence.				
ACCESSION	AQ437316				
VERSION	AQ437316.1	GI:4548655			
KEYWORDS	GSS.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 557)				
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and Hood,L.				
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome				
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)				
MEDLINE	99380589				
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources ( <a href="http://bacpac.med.buffalo.edu/ordering_bac.htm">http://bacpac.med.buffalo.edu/ordering_bac.htm</a> ) or from Resear h Genetics ( <a href="http://info@resgen.com">info@resgen.com</a> ). BAC end Web Server: <a href="http://www.htsc.washington.edu">http://www.htsc.washington.edu</a> Plate: 659 row: C column: 12 Seq primer: SP6 Class: BAC ends High quality sequence stop: 557. Location/Qualifiers 1..557 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="Plate=659 Col=12 Row=C" /clone_lib="RPCI-11 Human Male BAC Library" /sex="male" /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"				
BASE COUNT	227 a	127 c	108 g	79 t	16 others
ORIGIN					
Query Match		72.8%	Score 18.2;	DB 228;	Length 557;
Best Local Similarity		87.0%;	Pred. No. 2e+02;		
Matches	20;	Conservative	0;	Mismatches	3;
				Indels	0;
				Gaps	0;
QY	3	cattgtcgaagtccagacaatta 25			
Db	160	CAATGTCGCAGACACAGATCA 182			
RESULT	15				
AZ225990/c					
LOCUS	AZ225990	567 bp	DNA	GSS	14-JUN-2000
DEFINITION	RPCI-23-52D5.TV RPCI-23 Mus musculus genomic clone RPCI-23-52D5, DNA sequence.				
ACCESSION	AZ225990				
VERSION	AZ225990.1	GI:8534039			
KEYWORDS	GSS.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 567)					
REFERENCE					

The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

Email: mdadams@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html).  
Seq primer: M13 Reverse  
Class: BAC ends.

FEATURES Location/Qualifiers  
source 1..709

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2301C14"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelOBAC11; Site\_1: HindIII; Site\_2:  
HindIII"

BASE COUNT 303 a 109 c 124 g 173 t  
ORIGIN

Query Match 72.8%; Score 18.2; DB 223; Length 709;  
Best Local Similarity 87.0%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 acattgtcgcaagtcacagaatt 24

| ||||| ||||| ||||| |||||

Db 666 ATATTGTTGCAAGTAACAGAATT 644

RESULT 17

BE658717/c

LOCUS BE658717 751 bp mRNA EST 06-SEP-2000

DEFINITION GM700007A10F9 Gm-r1070 Glycine max cDNA clone Gm-r1070-2561 3',  
mRNA sequence.

ACCESSION BE658717

VERSION BE658717.1 GI:9984609

KEYWORDS EST.

SOURCE soybean.

ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.

REFERENCE 1 (bases 1 to 751)

Vodkin, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Corryell, V.,  
Erpelting, J., Raph, C., Shoop, E., Pardini, J., Liu, L. and Lewin, H.,  
A Functional Genomics Program for Soybean (NSF 9872565)

Unpublished (1999)

Other ESTs: AW156684 corresponding to Gm-cl015-2632 (5')

Contact: Vodkin, L.O., PI, A Functional Genomics Program for

Soybean (NSF 9872565)

Lewin, H. A., Director, Keck Center for Comparative and Functional  
Genomics

University of Illinois

Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA

Tel: (217) 244-6147

Fax: (217) 333-4582

Email: l-vodkin@uiuc.edu

This clone is available through: Genome Systems, Inc. 4633 World  
Parkway Circle St. Louis, Missouri 63134. For further information  
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)  
427-3324 or contact: clones@genomesystems.com or info@genome  
systems.com web site: www.genomesystems.com

Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES Location/Qualifiers  
source 1..751

/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="Gm-r1070-2561"  
/clone\_lib="Gm-r1070"

/note="The library Gm-r1070 is a sequence-driven, reracked  
set of 9,216 clones selected from cDNA libraries from  
various tissues and stages of development of soybean that  
represent 2,639 sequences from immature cotyledons, 1,770  
from immature seed coats, 3,938 from flowers, and 869  
from young pods. The 5' ESTs of the source clones form  
the different libraries was used to select singletons, or  
a representative of each contig, which were reracked to  
form library Gm-r1070. The cDNA clones of the reracked  
Gm-r1070 library were then sequenced at the 3' end. The  
contig analysis to select unique genes was performed by  
the laboratory of Ernest Retzel, Center for Computational  
Genomics and Bioinformatics, University of Minnesota,  
<http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html>  
. Reracking was performed by Genome Systems, St. Louis,  
<http://www.genomesystems.com>, and 3' sequencing by the  
Keck Center for Comparative and Functional Genomics,  
University of Illinois,  
<http://www.life.uiuc.edu/biotech/keck.html>. Note: The  
corresponding 5' EST from each clone in the Gm-r1070  
library is listed in the 'OTHER EST' field. The detailed  
information on the source library for each clone can also  
be obtained by referring to the Genome Systems clone ID of  
the original cDNA library that is also listed under  
'OTHER EST'."

BASE COUNT 193 a 171 c 131 g 226 t 30 others  
ORIGIN

Query Match 72.8%; Score 18.2; DB 138; Length 751;  
Best Local Similarity 87.0%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 catgtcgcaagtcacagaatta 25

| ||||| ||||| ||||| |||||

Db 118 CATACACGCAAGTCACAGAATTA 96

RESULT 18

AQ238325

LOCUS AQ238325

DEFINITION AQ238325 352 bp DNA GSS 21-APR-1999

DNA sequence.

ACCESSION AQ238325

VERSION AQ238325.1 GI:3670616

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 352)

Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,  
Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and Venter, J.C.

Use of human BAC End Sequences for Sequence-Ready Map Building

Unpublished (1998)

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from  
Research Genetics (info@resgen.com). BAC end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: n7

Class: BAC ends.  
Location/Qualifiers  
source 1..352

/organism="Homo sapiens"

```

/db_xref="GDB:7526293"
/db_xref="taxon:9606"
/clone="RPCI-11-69H14"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/notes="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"
BASE COUNT      86 a  91 c  98 g  75 t  2 others
ORIGIN

Query Match      71.2%; Score 17.8; DB 225; Length 352;
Best Local Similarity 90.5%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy  2 acattgtcgcaagtcacagaa 22
||||| ||||| ||||| ||||| |||||
Db  293 ACATGGTCTCAAGTCACAGAA 313

RESULT 19
R21156/c 379 bp mRNA EST 18-APR-1995
LOCUS yN20a04.r1 Soares placenta NB2HP Homo sapiens cDNA clone
DEFINITION IMAGE:130254 5', mRNA sequence.
ACCESSION R21156
VERSION R21156.1 GI:775937
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 379)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
,R., Williamson,A., Woldmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 987
High quality sequence stops: 337
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 987 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 337.
FEATURES
Location/Qualifiers
1. 379
/organism="Homo sapiens"
/db_xref="GDB:537603"
/db_xref="taxon:9606"
/clone="IMAGE:130254"
/clone_lib="Soares placenta NB2HP"
/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: placenta; Vector: p7T3D (pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'
AACTGAGAAATTCGCGCGCAGGAATTTTTTTTTTTTTTTT 3').
double-stranded cDNA was ligated to Eco RI adaptors
(pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo. "
```

```

BASE COUNT      123 a  56 c  83 g  117 t

Query Match      71.2%; Score 17.8; DB 187; Length 379;
Best Local Similarity 90.5%; Pred. No. 3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy  5 ttgtcgcaagtcacagaa 25
||||| ||||| ||||| ||||| |||||
Db  325 TTGTCCCAAGTCACACAATTA 305

RESULT 20
BE101686 382 bp mRNA EST 13-JUN-2000
LOCUS UI-R-B01-aps-a-08-0-UI.s1 UI-R-B01 Rattus norvegicus cDNA clone
DEFINITION UI-R-B01-aps-a-08-0-UI 3', mRNA sequence.
ACCESSION BE101686
VERSION BE101686.1 GI:8493784
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 382)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.wiueg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized mid-brain library cDNA library Preparation: M.B. Soares
Lab Clone distribution: clones will be available through Research
Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.
FEATURES
Location/Qualifiers
1. 382
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-B01-aps-a-08-0-UI"
/clone_lib="UI-R-B01"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The library
(UI-R-B01) is a subtracted library derived from a mixture
of the following tissues: thalamus, cerebellum,
hypothalamus, medulla, pons, midbrain, cerebral cortex,
corpus striatum and hippocampus. For a detailed
description of the library from which this clone was
derived, please visit our web site at
rategen.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LIB="UI-R-B01"
TAG_TISSUE="mid-brain"
TAG_SEQ="GATGG"
BASE COUNT      92 a  73 c  87 g  129 t  1 others
```



KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE  
JOURNAL  
COMMENT

FEATURES  
source

EST.  
Caenorhabditis elegans.  
Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea  
; Rhabditidae; Peloderinae; Caenorhabditis.  
1 (bases 1 to 225)  
Ward,S., Smith,H., Clifton,S., Marra,M., Hillier,L., Kucaba,T.,  
Pape,D., Martin,J., Wyllie,T., Underwood,K., Theising,B., Allen,M.,  
Bowers,Y., Person,B., Swaller,T., Steptoe,M., Gibbons,M., Harvey,N.,  
Ritter,E., Jackson,Y., McCann,R., Waterston,R. and Wilson,R.  
UofArizona-WashU C. elegans EST project  
Unpublished (1999)  
Contact: Samuel Ward, Ph.D.  
UofArizona-WashU C. elegans EST project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Contact Harold Smith (heseu.arizona.edu) for further information  
relating to organism, libraries, or clone availability.  
Trace considered overall poor quality  
Seq primer: -40up from Gibco  
High quality sequence stop: 1.  
Location/Qualifiers  
1. .225  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/clone\_lib="C elegans fem3 Q23 S1"  
/lab\_hosts="DH5alpha cells"  
/note="Vector: pBluescript II SK+; Site\_1: XhoI; Site\_2:  
NotI; This C. elegans library was made from fem3(q23)  
worms (produce only sperm at 25 C). cDNA was generated  
via oligo (dT) priming and directionally cloned into  
pBluescript II SK+ vector with a modified polylinker.  
This library was subtracted with a second fem-1(hc17)  
(produce only oocytes at 25 C) library to enrich this  
library for sperm specific genes. First strand cDNA  
synthesis was primed with a NotI-15t oligo (sequence  
5'-GACATGCTAGATCGAGCGGCCCTTTTCTTTT-3'). After  
2nd strand synthesis, XhoI-EcoRI adaptors (Stratagene  
#901120) were ligated, and cDNA was digested with NotI and  
ligated into XhoI-NotI digested vector."  
64 a 45 c 64 g 52 t

BASE COUNT  
ORIGIN

Query Match 70.4%; Score 17.6; DB 111; Length 225;  
Best Local Similarity 83.3%; Pred. No. 3.4e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 acattgtcgcaagtcacagaatta 25  
||||| ||||||||| ||| |||

Db 63 ACATTTTCGCAAGTCATTGAAGTA 40

RESULT 24  
AW833631  
LOCUS  
DEFINITION  
QV4-TT0008-161199-033-g05 TT0008 Homo sapiens cDNA, mRNA sequence.  
ACCESSION  
AW833631  
VERSION  
AW833631.1 GI:7927605  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 232)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

TITLE  
JOURNAL  
MEDLINE  
COMMENT

CONTACT: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gechtml2.pl?cl=st2-QV4-TT0008-161  
199-033-g05&t3=1999-11-16&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 232.  
Location/Qualifiers  
1. .232  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="TT0008"  
/dev\_stage="Adult"  
/note="Organ: testis; Vector: puc18; Site\_1: SmaI; Site\_2:  
SmaI; A mini-library was made by cloning products derived  
from ORESTES PCR (U.S. Letters Patent application No. 196  
716 - Ludwig Institute for Cancer Research) profiles  
into the puc 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."  
70 a 49 c 65 g 48 t

BASE COUNT  
ORIGIN

Query Match 70.4%; Score 17.6; DB 121; Length 232;  
Best Local Similarity 83.3%; Pred. No. 3.4e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaatt 24  
||||| ||||||||| |||||||

Db 59 GTCACCTGTTGCAAGTCCAGAAAT 82

RESULT 25  
AW833768  
LOCUS  
DEFINITION  
QV4-TT0008-130100-077-e06 TT0008 Homo sapiens cDNA, mRNA sequence.  
ACCESSION  
AW833768  
VERSION  
AW833768.1 GI:7927742  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 232)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

TITLE  
JOURNAL  
MEDLINE  
COMMENT

CONTACT: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922

FEATURES  
source

BASE COUNT  
ORIGIN

Qy 19

ORGANISM

## INTRODUCTION

## FEATURES

```

Query Match      70.4%; Score 17.6; DB 227; Length 403;
Best Local Similarity 83.3%; Pred. No. 3.7e+02; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 4;

QY 2 acattgtcgcaagtcacagaatta 25
    |||||  |||  |||||
Db 194 ACATTGCTAAATTACAGAATTA 171

RESULT 28
AQ314476      431 bp      DNA      GSS      04-MAY-1999
LOCUS      RPC111-104B8_TV RPC1-11 Homo sapiens genomic clone RPC1-11-104B8,
DEFINITION      DNA sequence.
ACCESSION      AQ314476
VERSION      AQ314476.1 GI:4045939
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 431)
AUTHORS      Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE      Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL      Unpublished (1998)
COMMENT      Other_GSSs: RPC111-104B8.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPC1-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buhalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buhalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.

FEATURES
    source
        Location/Qualifiers
            1..431
                /organism="Homo sapiens"
                /db_xref="GDB:7539583"
                /db_xref="taxon:9606"
                /clone="RPC1-11-104B8"
                /clone_lib="RPC1-11"
                /sex="Male"
                /cell_type="Lymphocytes"
                /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                    RPC111 Human Male BAC Library"

BASE COUNT      137 a      87 c      101 g      106 t
ORIGIN
    1..431
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="RPC1-11-104B8"
        /clone_lib="434 (synonym: htes3)"
        /tissue_type="testis"
        /dev_stage="adult"
        /lab_host="DH10B"
        /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"

BASE COUNT      159 a      69 c      125 g      107 t
ORIGIN
    1..460
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="DKFZp434H0915"
        /clone_lib="434 (synonym: htes3)"
        /tissue_type="testis"
        /dev_stage="adult"
        /lab_host="DH10B"
        /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"

Query Match      70.4%; Score 17.6; DB 105; Length 460;
Best Local Similarity 83.3%; Pred. No. 3.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaatt 24
    |||||  |||  |||  |||||
Db 226 GACATTGTAGCACATCATAGAAAT 203

RESULT 30
AQ142135/c    499 bp      DNA      GSS      24-SEP-1998
LOCUS      HS_3157_B1_D08_T7 CIT Approved Human Genomic Sperm Library D Homo
DEFINITION      sapiens genomic clone Plate-3157 Col=15 Row=H, DNA sequence.
ACCESSION      AQ142135
VERSION      AQ142135.1 GI:3532788
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 499)
AUTHORS      Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE      Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL      Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE      99380589
COMMENT      High Throughput Sequencing Center
Contact: Mahairas GG, Wallace JC, Hood L
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

```

Query Match	70.4%	Score 17.6;	DB 225;	Length 517;
Best Local Similarity	83.3%	Pred. No. 3.9e+02;		
Matches 20; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;





```

Query Match      70.4%; Score 17.6; DB 241; Length 609;
Best Local Similarity 83.3%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaatt 24
||||| ||||| ||||| ||||| |||||
Db 121 GACATTGTCCCAAGCCACATCATT 144

RESULT 35
AQ0602696
LOCUS      AQ0602696      615 bp      DNA      GSS      10-JUN-1999
DEFINITION HS_2116_A2_D07_T7C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2116 Col=14 Row=G, DNA sequence.
ACCESSION  AQ0602696
VERSION     AQ0602696.1 GI:5062772
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 615)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2050 row: M column: 3
            Class: BAC ends
            High quality sequence stop: 643.
            Location/Qualifiers
            1..615
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Plate=2050 Col=3 Row=M"
            /clone_lib="CIT Approved Human Genomic Sperm Library D"
            /sex="male"
            /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT    170 a 165 c 119 g 145 t 16 others
ORIGIN
1..615
acattgtcgcaagtcacagaatt 25
||||| ||||| ||||| ||||| |||||
Db 402 ACATTATGCAAGTCGCATAATTA 425

FEATURES
source
1..615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2116 Col=14 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

Query Match      70.4%; Score 17.6; DB 230; Length 615;
Best Local Similarity 83.3%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 acattgtcgcaagtcacagaatt 25
||||| ||||| ||||| ||||| |||||
Db 402 ACATTATGCAAGTCGCATAATTA 425

FEATURES
source
1..615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2116 Col=14 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

Query Match      70.4%; Score 17.6; DB 230; Length 615;
Best Local Similarity 83.3%; Pred. No. 3.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 acattgtcgcaagtcacagaatt 25
||||| ||||| ||||| ||||| |||||
Db 402 ACATTATGCAAGTCGCATAATTA 425

FEATURES
source
1..615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2116 Col=14 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

Query Match      70.4%; Score 17.6; DB 226; Length 643;
Best Local Similarity 83.3%; Pred. No. 4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 acattgtcgcaagtcacagaatt 25
||||| ||||| ||||| ||||| |||||
Db 438 ACATGTGCACCAAGTCAAGAATTA 415

FEATURES
source
1..643
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2050 Col=3 Row=M"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT    187 a 126 c 85 g 233 t 12 others
ORIGIN
1..643
acattgtcgcaagtcacagaatt 25
||||| ||||| ||||| ||||| |||||
Db 438 ACATGTGCACCAAGTCAAGAATTA 415

RESULT 37
AQ038199
LOCUS      AQ038199      659 bp      mRNA      EST      18-OCT-1999
DEFINITION EST279856 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA clone cLET1P15, mRNA sequence.
ACCESSION  AQ038199
VERSION     AQ038199.1 GI:5896953
KEYWORDS    EST.
SOURCE      tomato.
ORGANISM    Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
            1 (bases 1 to 659)
            D'Ascenzo,M., He,X., Lyman,J., Holt,I.E., Liang,F., Upton,J.,
            Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W.,
            Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni
            J.
            Generation of ESTs from tomato leaf tissue
            Unpublished (1999)
            Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU
            5 prime sequence.

```

```

ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 643)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE     99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2050 row: M column: 3
            Class: BAC ends
            High quality sequence stop: 643.
            Location/Qualifiers
            1..643
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Plate=2050 Col=3 Row=M"
            /clone_lib="CIT Approved Human Genomic Sperm Library D"
            /sex="male"
            /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT    187 a 126 c 85 g 233 t 12 others
ORIGIN
1..643
acattgtcgcaagtcacagaatt 25
||||| ||||| ||||| ||||| |||||
Db 438 ACATGTGCACCAAGTCAAGAATTA 415

Query Match      70.4%; Score 17.6; DB 226; Length 643;
Best Local Similarity 83.3%; Pred. No. 4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 acattgtcgcaagtcacagaatt 25
||||| ||||| ||||| ||||| |||||
Db 438 ACATGTGCACCAAGTCAAGAATTA 415

FEATURES
source
1..643
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2050 Col=3 Row=M"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT    187 a 126 c 85 g 233 t 12 others
ORIGIN
1..643
acattgtcgcaagtcacagaatt 25
||||| ||||| ||||| ||||| |||||
Db 438 ACATGTGCACCAAGTCAAGAATTA 415

RESULT 37
AQ038199
LOCUS      AQ038199      659 bp      mRNA      EST      18-OCT-1999
DEFINITION EST279856 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA clone cLET1P15, mRNA sequence.
ACCESSION  AQ038199
VERSION     AQ038199.1 GI:5896953
KEYWORDS    EST.
SOURCE      tomato.
ORGANISM    Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
            1 (bases 1 to 659)
            D'Ascenzo,M., He,X., Lyman,J., Holt,I.E., Liang,F., Upton,J.,
            Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W.,
            Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni
            J.
            Generation of ESTs from tomato leaf tissue
            Unpublished (1999)
            Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU
            5 prime sequence.

```

FEATURES	Location/Qualifiers
source	1. .659
	/organism="Lycopersicon esculentum"
	/cultivar="Rio Grande PTO"
	/db_xref="taxon:4081"
	/clone="cLETF15"
	/clone_lib="tomato mixed elicitor, BTI"
	/tissue_type="leaf"
	/dev_stage="4-6 week old plants"
	/lab_host="XJ1-Blue MRF"
	/note="Vector: pBlueScript SK(-); site_1: EcoRI; Site_2: XhoI; cLET - Inoculated with a variety of disease response elicitors. plants exposed to 2,6 dichloroisonicotinic acid, BTH, jasmonic acid, ethylene, fenthion, EIX, okadaic acid, or systemin prior to tissue harvest. EcoRI site was destroyed during cloning."
BASE COUNT	185 a 119 c 149 g 206 t
ORIGIN	
Query Match	70.48; Score 17.6; DB 110; Length 659;
Best Local Similarity	83.3%; Pred. No. 4e+02; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
QY 1	gacattgtcgcaagtccagaaatt 24
Db 421	GCATTGAAGCAGACAGAAATT 444
RESULT 38	
AQ389869/c	
LOCUS	AQ389869 661 bp DNA 21-MAY-1999
DEFINITION	RPC111-142A6-TV RPCI-11 Homo sapiens genomic clone RPCI-11-142A6, DNA sequence.
ACCESSION	AQ389869
VERSION	AQ389869.1 GI:4360892
KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1. (bases 1 to 661)
AUTHORS	Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C.
TITLE	Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL	Unpublished (1997)
COMMENT	Other_GSSs: RPC111-142A6.TJ Contact: Shaying Zhao, William Nierman, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850 Tel: 301 838 0200 Fax: 301 838 0208 Email: hbe@tigr.org Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources ( <a href="http://bacpac.med.buffalo.edu/ordering">http://bacpac.med.buffalo.edu/ordering</a> ) or from Research Genetics ( <a href="http://info@resgen.com">info@resgen.com</a> ). BAC end search page: <a href="http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html">http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html</a> Seq primer: T7 Class: BAC ends.
FEATURES	Location/Qualifiers
source	1. .661
	/organism="Homo sapiens"
	/db_xref="GDB:7554149"
	/db_xref="taxon:9606"
	/clone="RPCI-11-142A6"
	/clone_lib="RPCI-11"
	/sex="Male"
	/cell_type="Lymphocytes"
	/note="Vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI;

```

Db 180 ACATTGTCGATGCACAGAGTGA 157
||||| ||| ||||||| |
1 gacattgtcgaagtcacagaatt 24
||||| ||| | |||||||
Db 112 GACATTTTCACAGGCCACAGAATT 135

RESULT 41
LOCUS AQ0806020 531 bp DNA GSS 09-AUG-1999
DEFINITION HS_3237_A2_A03_MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3237 Col=6 Row=A, DNA sequence.
ACCESSION AQ0806020
VERSION AQ0806020.1 GI:5723352
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 531)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3237 row: A column: 6
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 531.
FEATURES
source
1..531
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3237 Col=6 Row=A"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 136 a 132 c 109 g 152 t 2 others
ORIGIN

Query Match 69.6%; Score 17.4; DB 233; Length 531;
Best Local Similarity 94.7%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 acattgtcgaagtcacag 20
||||| ||| ||||||| |||
Db 28 ACAGTGTCCGACAGTCACAG 10

RESULT 42
LOCUS AQ403329 612 bp DNA GSS 13-MAR-1999
DEFINITION HS_3061_B1_H12_T7A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=637 Col=23 Row=P, DNA sequence.
ACCESSION AQ403329
VERSION AQ403329.1 GI:4414109
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 612)

```

```

Db 180 ACATTGTCGATGCACAGAGTGA 157
||||| ||| ||||||| |
1 gacattgtcgaagtcacagaatt 24
||||| ||| | |||||||
Db 112 GACATTTTCACAGGCCACAGAATT 135

RESULT 40
LOCUS AZ334568 719 bp DNA GSS 29-SEP-2000
DEFINITION IM0064F02F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0064F02 F, DNA sequence.
ACCESSION AZ334568
VERSION AZ334568.1 GI:10402021
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 719)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0064 row: F column: 02
Seq primer: CGTTGTAACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 719.
FEATURES
source
1..719
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0064F02"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gil4732114|gb|AF129072.1), a copy number
inducible derivative of plasmid RI. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 230 a 111 c 108 g 270 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 241; Length 719;
Best Local Similarity 83.3%; Pred. No. 4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

**AUTHORS** Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

**TITLE** Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

**JOURNAL** Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

**MEDLINE** 99380589

**COMMENT** Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu  
plate: 637 row: P column: 23  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 612.

**FEATURES** Location/Qualifiers  
1..612  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="plate:637 Col=23 Row=P"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"  
BASE COUNT 165 a 132 c 113 g 171 t 31 others  
ORIGIN

Query Match 69.6%; Score 17.4; DB 228; Length 612;  
Best Local Similarity 94.7%; Pred. No. 4.9e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtccaca 19  
||||||| |||||||||  
Db 292 GACATTGTGGCAAGTCACA 274

**RESULT** 43  
A2385697/c  
LOCUS A2385697 622 bp DNA GSS 02-OCT-2000  
DEFINITION IM0144M15F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0144M15 F, DNA sequence.

ACCESSION A2385697  
VERSION A2385697.1 GI:10499397  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 622)  
**AUTHORS** Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

**TITLE** Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
**JOURNAL** Unpublished (2000)  
**COMMENT** Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0144 row: M column: 15  
Seq primer: CGTTGTAACAGCAGCCAGT  
Class: plasmid ends  
High quality sequence stop: 622.

**FEATURES** Location/Qualifiers  
1..622  
source

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0144M15"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides. The ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."  
BASE COUNT 201 a 128 c 126 g 167 t  
ORIGIN

Query Match 69.6%; Score 17.4; DB 242; Length 622;  
Best Local Similarity 94.7%; Pred. No. 4.9e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 acattgtcgcaagtccacag 20  
||||||| |||||||||  
Db 393 ACATTGTAGCAAGTCACAG 375

**RESULT** 44  
AV006577/c  
LOCUS AV006577 163 bp mRNA EST 25-AUG-1999  
DEFINITION AV006577 Mus musculus 18-day embryo C57BL/6J Mus musculus cDNA clone 1100001P13, mRNA sequence.

ACCESSION AV006577  
VERSION AV006577.1 GI:4783564  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 163)

**AUTHORS** Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K., Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Niihuma,H., Oda,H., Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y., Tomimaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.

**TITLE** RIKEN Mouse ESTs  
**JOURNAL** Unpublished (1999)

## COMMENT

Contact: Chie Owa  
Genome Science Laboratory  
RIKEN  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-9145  
Fax: 81-298-36-9098  
Email: genome-res@tc.riken.go.jp  
Thermotabilization and thermoactivation of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
Please visit our web site (http://genome rtc.riken.go.jp) for further details.

## FEATURES

source  
Location/Qualifiers  
1..163  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="1100001P13"  
/clone\_lib="Mus musculus 18-day embryo C57BL/6J"  
/sex="mixed"  
/dev\_stage="18-day embryo"  
30 a 35 c 34 g 60 t 4 others

BASE COUNT  
ORIGIN

Query Match 68.8%; Score 17.2; DB 108; Length 163;  
Best Local Similarity 86.4%; Pred. No. 5.1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 gacattgtcgcagtcacagaa 22  
||| ||||| ||| ||||| |||||  
Db 123 GAAATGTCATGTCACAGAA 102

RESULT 45  
AZ456884/c

LOCUS  
DEFINITION  
1M0260B02F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0260B02 F, DNA sequence. 04-OCT-2000

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 317)

REFERENCE  
AUTHORS

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts

TITLE

JOURNAL

COMMENT  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0260 row: B column: 02

Seq primer: CGTTGTAAACGACGCCAGT  
Class: plasmid ends

High quality sequence stop: 317.  
Location/Qualifiers

## FEATURES

source  
1..317  
/organism="Mus musculus"  
/strain="C57BL/6J"

/db\_xref="taxon:10090"  
/clone="UUGC1M0260B02"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PWD42 (gil4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 69 a 64 c 76 g 108 t  
ORIGIN

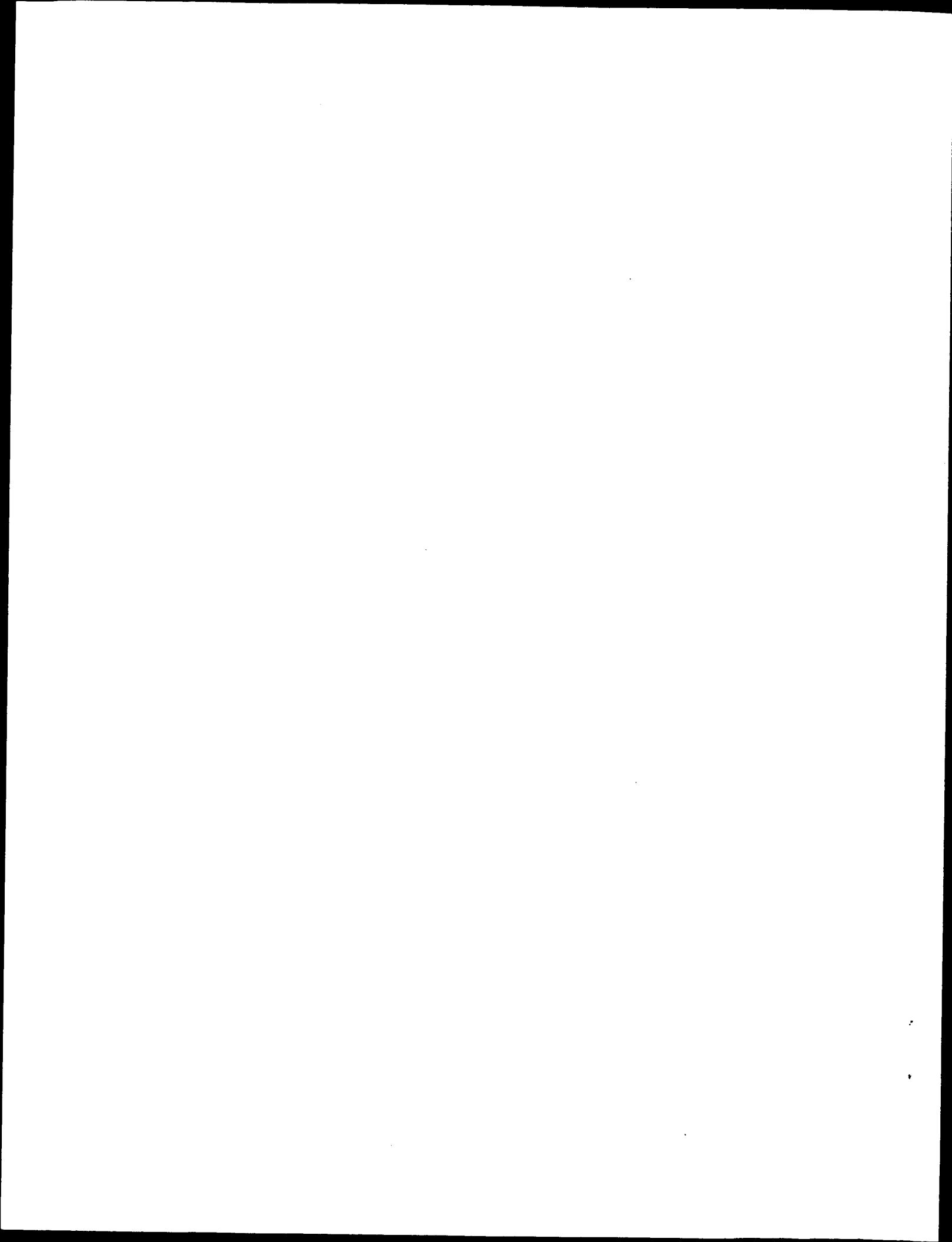
Query Match 68.8%; Score 17.2; DB 244; Length 317;  
Best Local Similarity 86.4%; Pred. No. 5.6e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 3 cattgtcgcagtcacagaaatt 24  
| ||||| ||||| ||||| |||||  
Db 66 CTTTGTACACAGTCACAGAACT 45

Search completed: October 9, 2001, 13:46:27  
Job time: 9502 sec

Wed Oct 10 07:44:19 2001

us-09-396-196f-1.std.rst

Page 23





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 12:14:40 ; Search time 1666.31 Seconds  
(without alignments)  
232.066 Million cell updates/sec

Title: US-09-396-196f-10  
Perfect score: 25  
Sequence: 1 tcgtcagggtcaggtcagcacgttg 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

##### GenEmbl :

```

1: gb_bal:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_bal:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

```

```

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_rod:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vil:*
59: gb_vil2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_rol1:*
95: gb_rol2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	25	100.0	1041	9	AR029499 Sequence
2	25	100.0	1041	9	AR034916 Sequence
3	25	100.0	1084	9	AI1530 BioB gene o
4	25	100.0	1121	10	E00893 Genomic DNA
5	25	100.0	5793	2	J04423 E.coli 7,8-
6	25	100.0	5872	9	A38246 Sequence 1
7	25	100.0	5872	9	A38251 Sequence 6
8	25	100.0	5872	9	A93674 Sequence 1

```

9 25 100.0 5872 9 A93679
10 25 100.0 5872 9 AR101809
11 25 100.0 5872 9 AR101810
12 25 100.0 11022 2 AE000180
13 23.4 93.6 15501 2 AF250776
14 23.4 93.6 13501 1 AE005258
15 23.4 93.6 297816 2 AF002553
16 22.4 89.6 7215 3 SNABIO
17 20.8 83.2 4000 2 AF250768
18 20.2 80.8 1041 3 EHB10TOP2
19 19.4 77.6 213732 1 AE001862
20 18.6 74.4 50195 65 AC019891
21 18.6 74.4 73076 4 AC003055
22 18.6 74.4 175455 70 AC026904
23 18.6 74.4 295312 4 AE003582
24 18.2 72.8 72289 93 HSDJ91115
25 18.2 72.8 72919 12 AC003096
26 18.2 72.8 83122 14 ATT20010
27 18.2 72.8 93330 67 AC022591
28 18.2 72.8 122951 89 AF312913
29 18.2 72.8 141990 85 AC004691
30 18.2 72.8 180257 60 AC007044
31 18.2 72.8 202828 72 AC064835
32 17.8 71.2 12891 1 AE004192
33 17.6 70.4 694 1 AF145866
34 17.6 70.4 1412 2 AF269004
35 17.6 70.4 4877 97 HUMHXA
36 17.6 70.4 7286 93 HSTNSCN
37 17.6 70.4 7346 9 AX022893
38 17.6 70.4 7346 9 AX030481
39 17.6 70.4 7390 97 HUMHXB
40 17.6 70.4 7560 9 AX045676
41 17.6 70.4 7560 93 HSTENAS3
42 17.6 70.4 11213 1 AE004950
43 17.6 70.4 19770 3 MFCI376
44 17.6 70.4 53533 2 BSY09476
45 17.6 70.4 72360 67 AC022984

ALIGNMENTS

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES Location/Qualifiers
source 1..1041
/organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.53; Mismatches 0; Indels 0; Gaps 0;
Matches 25; Conservative 0;

Qy 1 tcgtcaggtgcaggtcagcagcttg 25
|||||
Db 111 TCCTCAGGTGCAGGTGCAGCAGCTTG 135

RESULT 2
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES Location/Qualifiers
source 1..1041
/organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.53; Mismatches 0; Indels 0; Gaps 0;
Matches 25; Conservative 0;

Qy 1 tcgtcaggtgcaggtcagcagcttg 25
|||||
Db 111 TCCTCAGGTGCAGGTGCAGCAGCTTG 135

us-09-396-196f-10.std.rge

AR034916 AR034916 1041 bp DNA 29-SEP-1999
LOCUS AR034916
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES Location/Qualifiers
source 1..1041
/organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.53; Mismatches 0; Indels 0; Gaps 0;
Matches 25; Conservative 0;

Qy 1 tcgtcaggtgcaggtcagcagcttg 25
|||||
Db 111 TCCTCAGGTGCAGGTGCAGCAGCTTG 135

RESULT 3
LOCUS All530 1084 bp DNA 09-FEB-1994
DEFINITION BioB gene of E.coli with primers.
ACCESSION All530
VERSION All530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli
REFERENCE 1 (bases 1 to 1084)
AUTHORS
JOURNAL
FEATURES Location/Qualifiers
source Patent: GB 2216530-A 16 11-OCT-1989;
1..1084
/organism="Escherichia coli"
/db_xref="taxon:562"
24..1064
/gene="BioB"
24..1064
/gene="BioB"
/codon_start=1
/transl_table=11
/protein_id="CAA00965.1"
/db_xref="GI:490218"
/db_xref="SWISS-PROT:P12996"
/translation="MAHRPRWTLTSQVTELFEPKPLDLLFEAQVHRQHFDPQVQVST
LISIKTGACPEDCKYKPOSSRYKTEGLEARLMEVEQVLESARKAKAGSTRCMGAAW
KNPHERDMPYEQVQYKAMGLEACMTLGTLSAQARLANAGLDYYNHLDTSPF
YGNIIITRTYERLDTLEKVRDAGTKVCGSGTGVIGETVKDRAGLILQLANLPTPES
VFNNLVKVGKGPLADNDVDADFRTIAVARIMMPTSYVRLSAGRQMEQTOAMC
FMAGANSIFYGCKLILTPNPEDKDLQLFRKLGLNPQQTAVLAGDNEQQORLEQALMT
PDTDEYNAAL"
BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.52; Mismatches 0; Indels 0; Gaps 0;
Matches 25; Conservative 0;

Qy 1 tcgtcaggtgcaggtcagcagcttg 25
|||||

```





```

/evidence=experimental
/product="PROTEIN"
/protein_id="CAA02328.1"
/db_xref="GI:2294848"
/translation="MKLISDLRQDKLPHRHVFNMGYDGDNTSPHLAWDDVPAGTK
SFVVCYDPTGSGMHWVYVNLADPTVLPQFGSGVLVAMPDGLVQTRDFGKTK
YDGAAPRKGETHRYIFTVHALDIERIDVDEGASGMVGNVHFHSLASASITAMFS"
stem_loop
terminator
5583..5605
5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagcagcttg 25
|||||
Db 227 TCGTCAGGTGCAGGTGCAGCAGCTTG 251

RESULT 7
LOCUS A38251 5872 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 6 from Patent WO9408023.
ACCESSION A38251
VERSION A38251.1 GI:2294849
KEYWORDS Escherichia coli.
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 5872)
BIRCH.O., Brass,J., Fuhrmann,M. and Shaw,N.
BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
Patent: WO 9408023-A 6 14-APR-1994;
LONZA AG (CH)
Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 950809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
1141..1156
/clone="PBO30A15-9"
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAA02329.1"
/db_xref="GI:2294850"
/translation="MSWQEKINAAALDARRAALRRYPVAQAGRWLVADDFQYLN
SSNDYGLSHHPQIIRAWQQAQFGGSGSHVSGYSVVHQAEEELAEWLGYSRA
LLFISGPAANQAVIAAMAKEDRIAADRLSHASLLSPSLRRFAHNDVTHLAR
LLASPCPGQMVVTEGVFSMDGDSAPLAEIQVTTQCHNGWLMVYDDAHGTGVIGQGRG

```

```

SWLQKVKPELLIVTTEGKFGVSGAAVLCSSVTVDYLLQFARHLIYSTSMPPAQAAQAL
RASLAVIRSDGDRREKLAALITREFRAGVQDLPTFLADSCSAIPLIYVDNRSRALQL
AEKLRQGGQWVTAIRPTVPAGTARLRLTLTAHMQDIDRLLEVLHGNG"
3030..3045
/standard_name="BIOB RBS"
3043..3753
/gene="BIOB"
3043..3753
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYVTGCTDTEVGKTVASCALLOAKAAGRYTAGYKPVASGS
EKTPEGLRNSDALALORNSLQLDYATVNPYTFPAETSPHIIISAGEGRTIESLVMSAG
LRALQOADMVLVEGAGWFTPLSDTFTFADWVTPQLPVLVVGKLGCIINHAMLTA
QVIQHAGLTLAGVANDVTPPGKRHAEYMTTLTRMTPAPLLGELPWLAEIPNAATGK
YINLAFVASTLIGFTSRL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagcagcttg 25
|||||
Db 227 TCGTCAGGTGCAGGTGCAGCAGCTTG 251

RESULT 8
LOCUS A93674 5872 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent EP0798384.
ACCESSION A93674
VERSION A93674.1 GI:6741862
KEYWORDS Escherichia coli.
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 5872)
BIRCH.O. and Brass,J.
BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
Patent: EP 0798384-A 1 01-OCT-1997;
LONZA AG (CH)
Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A-15/9"
1..96
/function="PROMOTER PTAC"
/evidence=experimental
23..28
45..49
/standard_name="PROMOTER PTAC"
/evidence=experimental
105..119
/standard_name="BIOB RBS NO.9"
/evidence=experimental
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"
/codon_start=1

```

```

/transl_table=11
/evidence-experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHPRWTLTQVTELFKEPLDLLPEAQOVHROHFPDPROVQVST
LLSTKGPEDCKYQSSRYKTGLEAERLMEVEQVLESARKAKAGSTRFCMGAAW
KNPHERMDYQMVQVKGKLEACMTGLTLESQOARLANAGLDYNNHNLDTSPF
YGNITTTQYERLDITLEKVDAGIKVCSGTIGVGETVDRAGSLLLQLLANLTPPES
VPIMLVKVGTPLADNDVDADFIRTIIVARIIMPTSYVRLSAGREQNEQTOAMC
FMAGANSIFYGCKLLTTPNPEEDKDLQFLRKLGLNPOQTAVLAGDNEQQORLEQALMT
PDTDEYNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVYKQAIAAAFGRAAHYEQHADLQROSADALLAMLPRKYT
HVLDAQCPGWSHWRHRAQVLTALDSPMLVQARQKDAADHYLAGDIESLPLATA
TFDLAWSNLAVQWCGNLSALRELYRVVRPKGVVATFTLVQGSPLPHEQAWQVDERP
HANRPLPDELEQSLNGVHVHIOPIITLWFDLALSAMRSIKGIGATHLHGDRPRL
TRSQLRQLAWPOQQGKRYPLTYHLFLOVARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence-experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDDLAFDQRHWHPTYSMTSLPVPYVVSABGCELLISDGRR
LVDGSSWAAIHGVNHPQLNAAMKSIDAMSHVFGGITHAPAEILCRKILVAMTPOP
LECVPLADSGVAVVEMAKMALQYQAKGEARQRLFTFRNGYHGDFTGAMSVCDPDS
MHLWKYLPENLEFAPQPSRMDGEMDERDWGEARLMAAHRHEIAVILEPIVOGAG
GMRMTHPWLRIKICDREGILLIADIELATGRTGKLFACAEIAEAPDILCLGKAL
TGGMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLALIESGDWQ
QVADIEVQLREOLAPARDAEMVADVRLGAIQVVTTHPVNNAALQKFFVEQGVWIRP
FGKLYLMPPIYLLPQQLRLTAAVNRAVODETFEQ"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence-experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDGDKLPHRVFMGMGVDGDNISPHLAWDDVPAGTK
SFVVCYDPDAPTGSGWVWVNLNLPADTRVLPOGSGSLVAMPDGLVLOTDFGKIG
YDGAAPPKGETHRYIFTVHALDIERIDVDEGASAMGVNHFHSLASASITAMFS"
5383..5605
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
1318 a 1552 c 1695 g 1307 t

```

```

ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 5872;
Best local similarity 100.0%; Pred. No. 0.39;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgtcaggtagcaggtcagcagcttg 25
    |||||
Db 227 TCCTCAGGTGCGAGTCAGCACGTTG 251
    |||||

RESULT 9
LOCUS A93679 5872 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 6 from Patent EP0798384.
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 5872)
AUTHORS Birch, O. and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
source 1..5872
    /organism="Escherichia coli"
    /strain="DSM498"
    /db_xref="taxon:562"
    /clone="PBO30A15-9"
    1141..1156
    /standard_name="BIOF RBS"
    1154..2308
    /gene="BIOF"
    1154..2308
    /gene="BIOF"
    /standard_name="8-AMINO-7-OXONANOATE SYNTHASE"
    /EC_number="2.3.1.47"
    /codon_start=1
    /transl_table=11
    /evidence-experimental
    /product="KAPA SYNTHASE"
    /protein_id="CAB69594.1"
    /db_xref="GI:6741868"
    /translation="MSWQEK INAAIDARRAADLRRRYPVQAGRWLVADRDQVINF
    SNDYGLSHHPQILHAWQOQAFQFGSGSGSHVGSYVHVQALEEELAEWLYGSRA
    LFTISGFAANQAVIANMAKEDRIADRILSHASLLAEASLSPSOLRRFAHNDVTHLR
    LIASPCPGQOMVVTEGVFSMDGDSAPLAEIQVTOQHNGWLMVDDAHGTGVLGEQGRG
    SWLQVKPELLVVTGKFGVSGAALVCSVADYLLQFARHLIYSTSMPPAQAL
    RASLAVIRDEGDARKEALAILTRFAGVQDLPTFLADSCAIOPLIVGDNSRALQL
    AEKLRQOGCWTAIRPPTVPACTARLRLTLTAHEMODIDRLLEVLHGNG"
    3030..3045
    /standard_name="BIOD RBS"
    3043..3753
    /gene="BIOD"
    3043..3753
    /gene="BIOD"
    /standard_name="DETHIOBIOTIN SYNTHASE"
    /EC_number="6.3.3.3"
    /codon_start=1
    /transl_table=11
    /evidence-experimental
    /product="DTB SYNTHASE"
    /protein_id="CAB69595.1"
    /db_xref="GI:6741869"
    /translation="MSKRVFTCTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
    EKTPEGLRNSDALALQNSLQDLYATVNPYFAETPSHIIISAQGRPIESLVKSAG
    LRALEQGLADWLVVEGAGSFTPLSDTFTTFADWVTQBLPVILVGVKLGCIINHMLTA
    QVTHAGLTLAGVWANDVTPPGKRHAETMTLTRMIPAPILGEIPWLAENPENNAATGK

```

```

BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN
YINLAFVDASTLGFTSRL"
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagcacgttg 25
|||||
Db 227 TCCTCAGGTGCAGGTGCAGCACGTTG 251

RESULT 10
LOCUS      AR101809      5872 bp      DNA      PAT      14-FEB-2001
DEFINITION Sequence 1 from patent US 6083712.
ACCESSION AR101809
VERSION AR101809.1 GI:12812607
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
FEATURES
    source
        Location/Qualifiers
            1..5872
                /organism="unknown"

BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagcacgttg 25
|||||
Db 227 TCCTCAGGTGCAGGTGCAGCACGTTG 251

RESULT 11
LOCUS      AR101810      5872 bp      DNA      PAT      14-FEB-2001
DEFINITION Sequence 6 from patent US 6083712.
ACCESSION AR101810
VERSION AR101810.1 GI:12812608
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
FEATURES
    source
        Location/Qualifiers
            1..5872
                /organism="unknown"

BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagcacgttg 25
|||||
Db 227 TCCTCAGGTGCAGGTGCAGCACGTTG 251

RESULT 12
LOCUS      AE000180      11022 bp      DNA      BCT      01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete genome.
ACCESSION AE000180 U00096
VERSION AE000180.1 GI:1786988
KEYWORDS
SOURCE Escherichia coli K12.
ORGANISM Escherichia coli K12.
REFERENCE 1 (bases 1 to 11022)
AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (bases 1 to 11022)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 3 (bases 1 to 11022)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 4 (bases 1 to 11022)
AUTHORS Plunkett, G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT This sequence was determined by the E. coli Genome project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products: all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.
FEATURES
    source
        Location/Qualifiers
            1..11022
                /organism="Escherichia coli K12"
                /strain="K12"
                /sub_strain="MG1655"

```





biotin biosynthesis protein BioC (bioC) genes, complete cds; and  
dethiobiotin synthetase BioD (bioD) gene, partial cds.

AF250776  
AF250776.1 GI:12620124

uncultured bacterium pCosHE2.  
uncultured bacterium pCosHE2  
Bacteria; environmental samples.

1 (bases 1 to 5526)

Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streit, W.R.

Direct cloning from enrichment cultures, a reliable strategy for  
isolation of complete operons and genes from microbial consortia  
Appl. Environ. Microbiol. 67 (1), 89-99 (2001)

20575196  
11133432

2 (bases 1 to 5526)

Entcheva, P., Liebl, W. and Streit, W.R.

Direct Submission

Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet

Goettingen, Grisebachstr. 8, Goettingen 37077, Germany

Location/Qualifiers

1. 5526

/organism="uncultured bacterium pCosHE2"

/db\_xref="taxon:143797"

/clone="pCosHE2"

/note="unknown organism, cosmid clone derived from

environmental consortium"

complement(52..528)

/note="ORF1"

/codon\_start=1

/transl\_table=11

/product="hypothetical 17.1 kDa protein in modC-bioA

intergenic region"

/protein\_id="AAG60577.1"

/db\_xref="GI:12620125"

/translation="MKLISNDLRDQKLPKRRHVFNGMGVDGNISPHLAWDDVPACTK

SFVVTCDPAPGSGMHWVNLFPADTRVLQFGSGVLVAMPDGLQTRDFTGKAG

YDGAAPKGETHRIYFTHALDIERIDVDEGASGMVGFNVHFSIASASITAMFS"

complement(587..1876)

/gene="bioA"

complement(587..1876)

/gene="bioA"

/note="7,8-diaminopelargonic acid

synthetase-aminotransferase"

/codon\_start=1

/transl\_table=11

/product="DAPA-aminotransferase BioA"

/protein\_id="AAG60578.1"

/db\_xref="GI:12620126"

/translation="MTDDIAFDQRIHWPYTSMTSLPVPVVSAGCELISDGR

LVDMSSWAAIHGYNHPQLNAKMSQIDAMSHVFGGITTHAPATLCLKRLVAMTP

LCVFLADSGSVAEVMKMAQYQWAKGEARQRLTFRNGYHGTDFGASVCDPDS

MHSLWKGYLPENLFAPOSMDGWERDMYGFARLMAARHRETAATLIEIPVQAG

GRMYHPMLKTRKICDREGILLIADEIATGFGRTGKLPACAEHAIADPILCLGAL

TGSTMILSLTTREVAETISNGEACFMHGPFTFMGNPLACAAASLAILESQWQQ

QVADIEVQREQLAPDAEMADVRLVGAIGVYVTHPNNMAALQKFEVEQGVIRP

FGKLIYLPPIYILLPOOLRLTAANVRVQDTEFFCQ"

1963..3003

/gene="bioB"

/gene="bioB"

/codon\_start=1

/transl\_table=11

/product="biotin synthetase BioB"

/protein\_id="AAG60579.1"

/db\_xref="GI:12620127"

/translation="MAHRRRWTLISQVTELFKPLDLLEPAQQVIRHPDPRQVQVST

LLSIKTCAPEDCKYCPQSSRYKTGLEARLMEVQLSARKAKAAGSTFCMGAAW

KPNEDMDYLPQMGVQGLGLACMTGLTSESQAQRLANAGLLYNNHNDLTPDES

YGNITRYQERLDLKVVDAGIKVCSGGIVGLGETVKDRAGLLLOLANLPTTPES

VPNMLVKVGTPLANDDDVADFRTIATVARIMPTSYVRLSAGRCQMNQTOAMC

FMAGANSIFYGCKLLTTPNEEDKQLFRKLGKLPNQPTAVLAGNEQQORLEQALMT

PDDEYNNAAAL"

gene

/gene="bioF"

CDS

3000..4154

/gene="bioF"

/note="8-amino-7-oxononanoate synthase"

/codon\_start=1

/transl\_table=11

/product="KAPA synthetase BioF"

/protein\_id="AAG60580.1"

/db\_xref="GI:12620128"

/translation="MSWQBKINAAIDARRAADLRRYPVPAOGAGRWLVADROYLNF

SSNDYLGSLHHPQIIRAMQOAGAEQFCGSGSHVSGYSVHQALEEELAEWLGYSLA

LLFISGFAQAVIARMAKEDRIADRSLSHSLLEASLSFSQLRRFAHNDVTHLAR

LIASPCGQQLVVTEGVFSMDGDSAPLAEIQQVTOQHNGWLMVDAGHGTGVEQGRG

SCWLQKVPPELLVITFGKFGVGAALVLCSSVADYLLQFARHLIYSTMPPAQAL

RASLAVIRDEGDARREKLIVSLIAHFRAGVQDLPTFLADSCSAIQPLIVGDSRALQL

AEKLQCGCWTAIRPPTVPACTARLRLTLTAHEMQDIDRLLEVLHGNG"

4141..4896

/gene="bioC"

4141..4896

/gene="bioC"

/note="reaction step prior to synthesis of pimeloyl-CoA"

/codon\_start=1

/transl\_table=11

/product="biotin biosynthesis protein BioC"

/protein\_id="AAG60581.1"

/db\_xref="GI:12620129"

/translation="MATVKNQIAAFAAGRAALYEQHADLQKQSDVLLAMLPERKYT

HVLDAGCGFQWMTWRHRRHAQVLTALDLSPPMLVQARQKDAADHYLAGDIESPLATA

TFDLAWSLAVQWCGNLSTALRELYRVVRSGVAVFTTLVQSLPELHQAOAVDERP

HANRFLPDELEQSLNGVYOHIIQIPLTWEDDALSAMRSLKIGATHLHEGRDPRIL

TFSQIQLRLAMPQQQGRYPLTYHLFLGVARE"

4883..5526

/gene="bioD"

4883..5526

/gene="bioD"

/note="DTB synthetase"

/codon\_start=1

/transl\_table=11

/product="dethiobiotin synthetase BioD"

/protein\_id="AAG60582.1"

/db\_xref="GI:12620130"

/translation="MLVSKRYFVTGDTVEGKTVASCALLOAKAAGVYRTAGYKPVAS

GSEKTPGRLNSDALALORNSIOLDYATVNPYFAEPTSPHILSAQGRPIESVMS

SGLRALEQADWLVLEGAGGFTPLSDFTTADWVQQLPVILVVGKLCINHAML

TAAQLQHAGLTLAGVANDVTPPKRRAEYITTLTRMIPAPLIGEIPMLAENPE"

BASE COUNT 1274 a 1507 c 1567 g 1178 t

ORIGIN

Query Match

Best Local Similarity 93.6%; Score 23.4; DB 2; Length 5526;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tgcgcaggtgcaggtcagcagcttg 25

|||||

Db 2073 TCGTCAGTCCAGGTCCAGCAGCTTG 2097

RESULT 14

AE005258

LOCUS

DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82

of 155

ACCESSION AE005258 AE005174

VERSION AE005258.1 GI:12513751

KEYWORDS

SOURCE Escherichia coli O157:H7 EDL933.

ORGANISM

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

Escherichia.

REFERENCE 1 (bases 1 to 13501)

AUTHORS

Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,



```

ERGIVHSNRDLPVDPVAPPPELPSVDVYVNRSLVPGNVICKGNNAVYVEDADATKVLKM
PTTOSNEVTSVRCFNQYAGSAEKIYNGNGDILGIRMDXKINGSLNLSISLPAQ
AEHAYIDMFDRLEQKILFVDYTTETNVLYDRAKNEFNPIDISSNVSDRSWSQSQIMQ
SYHGKQDLISVQVLSKI"
gene
6462..7160
/ gene="20990"
CDS
6462..7160
/ gene="20990"
function="orf; Other or unknown (Phase or Prophage
Related)
/ note="No significant matches"
/ codon_start=1
/ transl_table=11
/ product="unknown protein encoded by prophage CP-933K"
/ protein_id="AAG55143.1"
/ db_xref="GI:12513758"
/ translation="MRPTSLNLVHQSSRSSMSTDTIESLVKASSVQWIKNNPQLRF
QCTDHNITQQIEAALDKTSTGTVLLNAIESIRLKSSETVVIHLSSRLQPMHARD
IDAEHRTGSDFCNNAVEYPCGEGISVVDYFATVFFHELLHVFNLNCRUKVES
SRAESKYSPLLEEARVTGLGAFSEEVLSENKFEEIGMPRTSYFSDSALIHDDNT"
VSLGQQVRLHPLI"
gene
complement(7668..8144)
/ gene="ybhB"
/ note="Z0992"
CDS
complement(7668..8144)
/ gene="ybhB"
function="orf; Unknown function"
/ note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MGL655: B0773"
/ codon_start=1
/ transl_table=11
/ product="orf, hypothetical protein"
/ protein_id="AAG55144.1"
/ db_xref="GI:12513759"
/ translation="MKLTISNDLRDGDKLPHRVFNGMGYDGDNIISPFLAWDDVPAGTK
SFVTCYDPDAPTSQGMWHVYVNLNLPADTVLPQFGSLVAMPDGVLTQTFDQKTG
YDGAAPPKGETHRIETVHALDVERIDVDEGASCAMGVNFVHFHSLASASITAMFS"
complement(8203..9492)
gene
/ gene="bioA"
/ note="Z0993"
CDS
complement(8203..9492)

Query Match          93.6%; Score 23.4; DB 1; Length 13501;
Best Local Similarity 96.0%; Pred. No. 1.8;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtgcagcacttg 25
|||||
Db 9689 TCGTCAGGTGCAGGTGCAGCACCCTTG 9713

RESULT 15
AP002553
LOCUS 297816 bp DNA BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION AP002553 BA000007
VERSION AP002553.1 GI:13360211
KEYWORDS
SOURCE
Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (sites)
AUTHORS Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C. and Shinagawa,H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)

MEDLINE
REFERENCE
AUTHORS Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
TITLE Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MGL655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
MEDLINE
REFERENCE
AUTHORS Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Shinagawa,H.
TITLE Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
MEDLINE
REFERENCE
AUTHORS Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
TITLE Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
MEDLINE
REFERENCE
AUTHORS Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
Hayashi,T.
TITLE Direct Submission
JOURNAL Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
URL:http://www.gen-info.osaka-u.ac.jp/, tel:81-6-6879-8365,
Fax:81-6-6879-2047)
genome project.
FEATURES
source
1..297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
79..1245
/ gene="ECs0753"
79..1245
/ gene="ECs0753"
/ note="similar to SUCC_EC01 gi11786948 percent identity
100 in 388 aa (Conserved in E.coli K-12)"
/ codon_start=1
/ transl_table=11
/ evidence="not experimental"
/ product="succinyl-CoA synthetase beta subunit"
/ protein_id="BAB34176.1"
/ db_xref="GI:13360212"
/ translation="MNLHEYQAKQLFARYLPAPGYACTTPREAEAAKSTGAGPW
VKQVHAGGRGKAGGVKVVNSKEDIRAFENMLGRKLVYQTDANGQPVNQILVEAAT
DIAKELYLGAVDVRSSRVFNAETEGVEIEKVAETPHLTHKVALDPLTGMPYQG
RELAFKLGLEGLVQOFTKIFMGLATIFLERDLALTEINPLVITKOGDLICDGLGA
DGNLFRQDPLREMRQOSEDPREAQAQWELNYVALDNGICMWNAGLAGMTWDIV
KLHGGEFANFLDVGGAATKERVTEAFKIILSDDKVAKVLNIFGGIVRCDLIADTIG
AAVEGVNVPVVRVLEGNNAELGAKKLADSLNITAAKGLTDAAGQVAAVBEK"
1245..2114
/ gene="ECs0754"
1245..2114
/ gene="ECs0754"
/ note="similar to SUCD_EC01 gi11786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/ codon_start=1
/ transl_table=11
gene
CDS

```



## COMMENT

Submitted (25-Aug-1993) to DDBJ by:  
Naoki Sakurai  
Res. Lab. of Applied Biochemistry  
Tanabe Selyaku Co., Ltd.  
2-50 Kawagishi-2-chome  
Toda, Saitama 335  
Japan  
Phone: 048-433-2545  
Email: nsakurai@dbj.nig.ac.jp  
Fax: 048-433-2540.

## FEATURES

## source

1. .7215

/organism="Serratia marcescens"  
/strain="Sr41"  
/db\_xref="taxon:615"  
complement(1353. .2630)  
/gene="bioA"  
complement(1353. .2630)  
/gene="bioA"  
/EC.number="2.6.1.62"  
/codon\_start=1

/product="7,8-diamino-pelargonic acid aminotransferase  
(DAPA aminotransferase)"

/protein\_id="BAA04284.1"  
/db\_xref="GI:402531"  
/translation="MSVTSADLAFDQRIHWPYTSMSRPLPCYPPIESASGVELQLADG  
RSIVDGMSSWMAIIGHYHNPRLNQAASQQLKMSHVMFGGITHPAALISLCRLVEMTP  
EALQCVFLADSGVAVESKMAIQYQARGERRQRIILLRHGYHGDITFGAMVSCDPD  
NSMSLQYQGLAPHLFATPCQRFDEWREEDIAFPAALEHQHAGEVAAYVILEPVPVQ  
AGMWRIYPTLVKRVDRVVAIKLLIADIEIATGRTGKLFACFAHOVVPDILCIG  
KALGGVYLSATLTTHVAETISNGAAGCFMGPTFMGNPLACAVADASLALLAENR  
WQAVSAETQKRELLPLGCRWRPHRVIGALGVVEMREPVDVAELQRFVVERGVW  
IRPGKLIYLPMPYIIIEAQLSLRTAAVADAAR"

## gene

2716. .3756

## CDS

2716. .3756

/gene="bioB"

/gene="bioB"

/codon\_start=1

/translation="biotin synthetase"

/protein\_id="BAA04285.1"

/db\_xref="GI:402532"

/translation="MMADRIHTWVQQAQLFKDPLLELLEFAQTVHQRHEDPRQVQVS  
TLISIKGACPECKYCPSPRYKTGLESERLMQVEQVLESARRAKANGSTRFCMGAA  
WKNPHRDMPYLOQMVGKAMGEMTCMTGLTLDGTOAERLAEAGLDYYNNHLDTSPE  
FYGIIITRSTYQERLDTLVKRDAGIKVCSGGIVGLGETVRDAGLLVQLANLPKPE  
SVPIIMLVKSGTPLADNDVDPDFRTIATAVRIMPPSSVYRLSAGREOMNEOTQAM  
CFMAGANSIFVGCKLLTTPNEEDKDLQLFKRLGLNPQQTATEHGDNQOQOVLAKQLL  
NADPAEYNAP"

## gene

3756. .4904

## CDS

3756. .4904

/gene="bioF"

/gene="bioF"

/codon\_start=1

/product="7-keto-8-aminopelargonic acid synthetase (KAPA  
synthetase)"

/protein\_id="BAA04286.1"

/db\_xref="GI:402533"

/translation="MSWQRIEQALAERLNAAVRRQTTGCGNRQIRLGDRLYLNLF  
SGNDYLGSDQARVIAWQARGYGVNGSGHVGFSAAHQALEGLAALWLGTPRAL  
LFITYANQAVLAALMQKDRILADRLSHASLLEAGAVAGVRRAPPAQSTAGLARIC  
LAKPCDQRLAVTBGLFSMDGDAPLAEHLRLTRAAGAWLMDVDAHGIGYRGEQGRGS  
CWQOQRPILLVAFGAFGVSGAAVLCDEATRYLQFARHLIYSTAMPAPACALQ  
AALARIREDLRLQDNRFRQGAAPLALITDSDTAIQPLLVGDNRQALDLATR  
IREGLGWSAIPPTVPPGGRNLRLITTAHQSQDIDRLLEVLNDVVSQ"

## gene

4888. .5655

## CDS

4888. .5655

/gene="bioC"

/gene="bioC"

/function="synthesis of pimeloyl-CoA"

/codon\_start=1

## gene

5648. .6331

## CDS

5648. .6331

/gene="bioD"

/gene="bioD"

/function="conversion from 7,8-diaminopelargonic acid into  
dethiobiotin"

/codon\_start=1

/translation="1"

/product="dethiobiotin synthetase (DTB synthetase)"

/protein\_id="BAA04288.1"

/db\_xref="GI:912453"

/translation="MSKRFVVTCTDEYKSVASSALLQANRACYPSPAGYKPVASGS  
EMTAGLRNGDALQANSVGLDYDEVNYPVFAEPTSPHIVSADGPRIDARLSDG  
LRRLQRADWLVEGAGWFTPLSAEYTFADWVROQLPVILVVGKILGICINHVLTA  
QAVQOAGLTLAGWIANVGGAPRRHQEYLATLRMLPRCROWAKSRTCRPERAPLQG  
YLDISLQAQ"

BASE COUNT 1347 a 2225 c 2367 g 1276 t

ORIGIN

Query Match 89.6%; Score 22.4; DB 3; Length 7215;

Best Local Similarity 95.8%; Pred. No. 5.6;

Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 cgtcaggtcgagtcgacgacgttg 25

|||||

Db 2830 CGTCAGGTGCAGGTCCAGCAGCTG 2853

|||||

RESULT 17

AF250768

LOCUS

DEFINITION

AF250768 4000 bp DNA BCT 31-JAN-2001

Uncultured bacterium pCosFS1

cds; BioB-like protein (bioB) and KAPA-synthetase (bioF) genes,

complete cds; and biotin biosynthesis protein BioC (bioC) gene,

partial cds.

AF250768

AF250768.1 GI:12620095

uncultured bacterium pCosFS1.

uncultured bacterium pCosFS1

Bacteria; environmental samples.

1 (bases 1 to 4000)

Entcheva, P., Liebl, W., and Streit, W.R.

Direct cloning from enrichment cultures, a reliable strategy for

isolation of complete operons and genes from microbial consortia

Appl. Environ. Microbiol. 67 (1), 89-99 (2001)

20575196

MEDLINE

PUBMED

11133432

REFERENCE

2 (bases 1 to 4000)

Entcheva, P., Liebl, W., and Streit, W.R.

Direct Submission

Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet

Goettingen, Grisebachstr. 8, Goettingen 37077, Germany

Location/Qualifiers

1. .4000

/organism="uncultured bacterium pCosFS1"

/db\_xref="taxon:143795"

/clone="pCosFS1"

/note="unknown organism, cosmid clone derived from

environmental consortium"

complement(<1. .1177)

/gene="bioA"

complement(<1. .1177)

gene

CDS

```

/gene="bioA"
/codon_start=1
/transl_table=11
/product="BioA-like protein"
/protein_id="AAG60558.1"
/db_xref="GI:12620098"
/translation="MSVTDSDLEFQDRIWHPTSMHPLPCYPVEAASGVLEQLADG
RRLIDGSSWAAITGYNHNPOLNLAATQLEKMSHVMFGGITHPAANELCRKYVAMP
EALOCVFLADSGSVAVESLKMALQYQWQGERRQRIILTLRHGYHGDTFGAMSVCDPD
NSMSLQGYLAPHLFAFAPQCQGDQWENDIAPFAALLEHQHAGEVAAILPEVVOG
AGGRIHYPTLRYLRDLQYLLIADETATGFGRTGKLFACEHAENVYDITCLGK
ALTGGYTLTLTRHVAETISNGAACFMHGFTEGNPLACAVANASLSLAENRW
QQQVAAITQOLKQOLLPTNPVEDKDLQFLKRLNPPQQTETEHGDNQQQVEVLAARIM
QADTAEFYNAL"
2305..2305
/gene="bioB"
/translation="MMADRIHWTVGLAQTLFDPKLELLEFEAQTVHRHQHDPDQVQS
TLLSIKTCACPEDCKYCPQSSRYKTGLESLRMQVLESAKAKAGSTRFCMGAA
WKNPHERDMPYLOQWQGVKAMGMETCTGLTLDGSAERLAEGLDYNNHNLDTPE
FYGNITITRSYQERLDTLQKRGAGIKVCSGGIVGLGETVXXRAGLLVQLANLPTPE
SVPINMLVKYKGTPLADNDVDVDFPFIITAVARIMPPSSVRLSAGREOMNEQTOAM
CFMAGANSIFYGCKLLTTPNEEDKDLQFLKRLNPPQQTETEHGDNQQQVEVLAARIM
QADTAEFYNAL"
2305..3456
/gene="bioB"
2305..3456
/gene="bioB"
/translation="7-keto-8-aminopelargonic acid synthetase; similar
to BioB"
/codon_start=1
/transl_table=11
/product="KAPA-synthetase"
/protein_id="AAG60556.1"
/db_xref="GI:12620096"
/translation="MSWQRIQEALEERXOSAARHRLVIKGGNGRYLQNDORYLNF
SGNDYLGLSQNAOVIAAWQGAQYGVSGSGHVTGCTAHOALQQLAQLGYPRS
LLFTSYAANQALLAALMOGEDRLADRLSHALLEAAHSPAQLRFRHNPQALAD
LLKXPCSQOILAVTEGVFSDGSAPELRLTRDAGAWLMDVDDHAGIGVQGEQGRG
SCQWQVRELLVYTFKAGFLGSAVLCDSEAEYLLQFARHLIYSTAMPQAQAL
QAALHCLOQGDLELRLNIQRFROGAQLSLTLESFAIQIQLVGNQRALDLAQ
HLRQGLWVSAIRPPTVPPGGARLITLTAHQPDIDRLLEVLHDVSV"
3440..>4000
/gene="bioC"
3440..>4000
/gene="bioC"
/translation="MTSAFEQVKNQAVASAFSRAAVGYDAAVQLRDVGERLLGMQGE
HQGQLLDACGGTGFYSRRWRELKQVLTALDLPAGMLAFARQQAADHYLLGDIENVP
LGDAAVDFSSLLVWQVCSLDLPRALAEIYRVRPGGVILFTSLAEGSLHGLGDAWQOV
DGERHVNDFLPLAQIEAACAGYRHLLQ"
794 a 1164 c 1236 g 802 t 4 others
BASE COUNT
ORIGIN

Query Match 83.2%; Score 20.8; DB 2; Length 4000;
Best Local Similarity 91.7%; Pred. No. 33;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 cgtcaggtgcaggcagcagcttg 25
|||||

```

```

Db 1379 CGTCAGGTGAGGTGAGCAGCTG 1402

RESULT 18
LOCUS EHB10TOP2 1041 bp DNA BCT 04-NOV-1996
DEFINITION Erwinia herbicola biotin synthetase (bioB) gene, complete cds.
ACCESSION U38648
VERSION U38648.1 GI:1228112
KEYWORDS 2 of 2
SEGMENT
SOURCE
ORGANISM
Erwinia herbicola strain=Eh010.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Pantoea.
1 (bases 1 to 1041)
Wu,C.H., Chen,H.Y. and Shiuan,D.
Isolation and characterization of the Erwinia herbicola bio operon
and the sequences of the bioA and bioB genes
Gene 174 (2), 251-258 (1996)
JOURNAL
MEDLINE 97045821
REFERENCE 2 (bases 1 to 1041)
AUTHORS Shiuan,D.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-1995) David Shiuan. Biology, Natl. Sun Yat-Sen
University, Kaohsiung, Taiwan, ROC
FEATURES
Source
1..1041
/organism="Pantoea agglomerans"
/strain="Eh010"
/db_xref="taxon:549"
1..1041
/gene="bioB"
1..1041
/gene="bioB"
/codon_start=1
/transl_table=11
/product="biotin synthetase"
/protein_id="AAC44534.1"
/db_xref="GI:1228114"
/translation="WAHARWTLISOVTELFDPKPLDLLFEACTIHRQHFDPQVQVST
LLSIKTCACPEDCKYCPQSSRYKTGLESLRMQVLESAKAKAGSTRFCMGAAW
KNPHERDMPYLOQWQGVKAMGLEACMTGLTDETAQORLASAGLDYNNHNLDTSPF
YGNITITRYQERLDTLQKRGAGIKVCSGGIVGLGETVKKRAGLLQLANLPTTPES
VFINMLVKYKGTPLADNDVDVDFPFIITAVARIMMPTSFVRLSAGREOMNEQTOAMC
FMAGANSIFYGCKLLTTPNEEDKDLQFLKRLNPPQQTETEHGDNQQQKLEQIFN
ADTDQFYNAAL"
230 a 322 c 308 g 181 t
BASE COUNT
ORIGIN

Query Match 80.8%; Score 20.2; DB 3; Length 1041;
Best Local Similarity 88.0%; Pred. No. 79;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggcagcagcttg 25
|||||
Db 111 TCCTCAGGTTCAGGTGAGCAGCTG 135

RESULT 19
LOCUS AE001862/c 213732 bp DNA BCT 22-NOV-1999
DEFINITION Deinococcus radiodurans R1 section 1 of 2 of the complete
chromosome 2.
ACCESSION AE001862 AE001825
VERSION AE001862.1 GI:6460468
KEYWORDS
SOURCE
ORGANISM
Deinococcus radiodurans.
Deinococcus radiodurans
Bacteria; Thermus/deinococcus group; Deinococcales; Deinococcus.
REFERENCE 1 (bases 1 to 213732)
AUTHORS White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,

```

Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,  
 Moffat, K.S., Qian, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,  
 Yamathavan, J.J., Lam, P., McDonald, L.J., Utterback, T., Zalewski, C.,  
 Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.  
 Genome sequence of the radioresistant bacterium *Deinococcus*  
*radiodurans* R1  
 Science 286 (5444), 1571-1577 (1999)  
 20036896  
 2 (bases 1 to 213732)  
 White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,  
 Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,  
 Moffat, K.S., Qian, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,  
 Yamathavan, J.J., Lam, P., McDonald, L.J., Utterback, T., Zalewski, C.,  
 Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W.,  
 Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S.,  
 Smith, H.O., Venter, J.C. and Fraser, C.M.  
 Direct Submission  
 Submitted (02-NOV-1999) The Institute for Genomic Research, 9712  
 Medical Center Dr, Rockville, MD 20850, USA  
 FEATURES  
 Location/Qualifiers  
 1..213732  
 /organism="Deinococcus radiodurans"  
 /strain="R1"  
 /db\_xref="taxon:1299"  
 /chromosome="2"  
 653..1435  
 /gene="DRA0001"  
 653..1435  
 /gene="DRA0001"  
 /note="similar to GB:D26185 SP:P37522 PID:467381  
 PID:580906 GB:AL009126 percent identity: 54.96; identified  
 by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="chromosome partitioning ATPase, putative, ParA  
 family"  
 /protein\_id="AAF12301.1"  
 /db\_xref="GI:6460595"  
 /translation="MVSARKTLTVFNHAGGAKTSLTNVGYELARGRLVLLDLDP  
 QANLWGLIGISGVTRMTPVAVDQPLSPVAKFGLDVPVHVSIAVAGQMMGRV  
 GAQRLRAEASGDYDVALIDSPSLGQILAILAALADQMTVPVPTROKGLDALPG  
 LOGALTEYREVRLPTVALYVPTFYDARRRHQDEVLADLKAHLSPLARPVPQREAVL  
 DSTAQAQPVSEYAPGTPVHADVQRLTADIAAIGVAYGENA"  
 1432..2313  
 /gene="DRA0002"  
 1432..2313  
 /gene="DRA0002"  
 /note="similar to GB:I34077 GB:X85964 PID:407375  
 PID:757761 PID:I163135 percent identity: 57.34; identified  
 by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="chromosome partitioning protein, ParB family"  
 /protein\_id="AAF12300.1"  
 /db\_xref="GI:6460594"  
 /translation="MTRRRPERRDLCLLGETPVDLSQANDIRALPVNELKVGSTQP  
 RSPFLRSELSIRAHGVQLPLLVRSDVQVEIVAGERRAAQLAGLAEVPPVV  
 RLSNEQRAALINENQDLNVIDVDGKLELTALTGLEREAKRLMLLRVVP  
 GDEHQLDQVFRSMGETWRTFAKNKRLINLPQVPLEALRAGLPLTIGSVVASPPER  
 QAEILLQAQNGASRSQQLQTPQTSVAVPEHFVAKVLSKRFSLGSLDTPTREALDR  
 WLAMPVRVRAIDFOS"  
 complement(2418..3881)  
 /gene="DRA0003"  
 complement(2418..3881)  
 /gene="DRA0003"  
 /note="similar to GB:M88334 SP:P25526 PID:147901 GB:U00096  
 PID:1789015 percent identity: 69.60; identified by  
 sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="succinate-semialdehyde dehydrogenase"  
 /protein\_id="AAF12294.1"  
 /db\_xref="GI:6460588"

/translation="MTTDLRTTVSSVTRSQAYFDGEWRNAPENFVRHPGNCVEIGE  
 VADCTPDARQAIADAEVALRENRQVNPVERGKILLRWHDLMFEHKEELAQLMTLEMG  
 KPISETGEVHYAAASFTEWCAEEAGRIAGERINLRFPKRGKLTISPEVGIYAVTTPWN  
 FPAGMITRKAACAAACMLPAELSPMTALYLTLMKLAGGPANTIVQVLPNDAS  
 ALTPQPMNDSRVKLTFTSTGEVGLTQQAGTIKRVSLKLEGGHAPFLVFDADLER  
 AASEVVASKFSNSQTCVCTNRVYVQGVAAEFIRLTETKTAALQIQDGFDEATQVGP  
 VFEQAGLDKVOQVODALTKGAOATGGVSSGLTFQPTVLVDVAPDSLLIRETFGP  
 VAPVTFDTEEGRLANDSEYGLAAYATRDIGRAFRIAEGLEYIGYVINGINDGLPSSA  
 APIHVPFGMKNSVGRGGHGWGLEYLETKFVSLGUS"  
 4024..4668  
 /gene="DRA0004"  
 4024..4668  
 /gene="DRA0004"  
 /note="similar to GP:2924378 percent identity: 55.29;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAF12299.1"  
 /db\_xref="GI:6460593"  
 /translation="MNSLSLAALLMAGSTRMGQPKALLPLGGMPLCRHAAQTIVE  
 AGYEVMAVLPGETGDVRAALQGLPLRWVTNPQALGLSFSFQAGGALWAEASAP  
 AVVFTLADMPLRDTHQALGTVFGEAQPAAIITRYGEVAAAPTLLRLRLLPRLEHFP  
 PSDHGPRTLRLGRLKKTDRPAELLIDITPEAAQAQQLFAEVTPGAQDGP"  
 complement(4719..5897)  
 /gene="DRA0005"  
 complement(4719..5897)  
 /gene="DRA0005"  
 /note="similar to GB:L33464 SP:P47734 PID:496118 percent  
 identity: 74.23; identified by sequence similarity;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="alcohol dehydrogenase, zinc-containing"  
 /protein\_id="AAF12293.1"  
 /db\_xref="GI:6460587"  
 /translation="MKALVQGVNKVGVETVPDPTLLPTDAIVKITSTAICGSDLHL  
 LDGYTSMETGDIILGHEFMGEVSEVRKLVKGVRRVVPFNIAGVCDPCPKRFES  
 KDSQVNLPHHMAALYAGTGSGALGYSHMYGGYAGGAAQYVRVPADVGPFKIESGL  
 KDFVLTFTDFTPTGQAQCNIVPGRDVAVVAGVQGFQAASQMLGAHVIV  
 DYPERLMAEAACQCTINYEKEDVLSLEATGGRGPHVIDAVGMEAHGHPGSTM  
 DKYQNLKLTFRITLRLWAILSCAKGGTVMFVYGLVDMKMPICAAAKGLIFRMG  
 QTHRHIPSLLLGRIAGQIDPSFVITHRASLDQAPELYKTFRDKQDCCLIKVVLNFWA  
 "  
 complement(5894..6700)  
 /gene="DRA0006"  
 complement(5894..6700)  
 /gene="DRA0006"  
 /note="similar to PID:1653488 percent identity: 48.46;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAF12292.1"  
 /db\_xref="GI:6460586"  
 /translation="MRRSSPQLRTWNNREQAPAPHAGATGPNSSFGATMQPLERSL  
 TGAVGAALAAAGLRSGLEKVVLTGLAGLIAVAAQGNPLATLRIKQNDGGTIV  
 RDAVTGKPAEQIYAVWRDPLGLPLMLHLRSVEVLDDKRSWTVAPALPGVSWEA  
 ELTADEPKRIAWRSIPGARIENSGVLFRRPAPGARGTEVVRLTVPFGGSGAGVIA  
 RMFNOEFSQQLRDLMLRFKREQLGHAPTAAQGGSSGAAALTPDNKRGLV"  
 6839..7744  
 /gene="DRA0007"  
 6839..7744  
 /gene="DRA0007"  
 /note="similar to GB:AL009126 percent identity: 62.98;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ABC transporter, ATP-binding protein, putative"  
 /protein\_id="AAF12291.1"  
 /db\_xref="GI:6460585"  
 /translation="MITTFQVSKTYGHVTALSDFNLTLTGELTALLGPNAGKSTA  
 IGLLLGLSAPSAGQVRLGADPRNDVRARIGAMPQESALPAGLTVREAVTLFASFP





2\_c10 (AC002948), 2\_g9 (AC002953), 1\_g7 (AC001990), 1\_b5 (AC001989), 2\_c3 (AC002949), 2\_h2, 2\_c4 (AC002950), 2\_b6.

FEATURES  
 source  
 1. .73076  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /chromosome="2L"  
 /map="22F2-22F4"  
 /clone="P1 DS06332 (D91)"  
 21203 a 14890 c 15769 g 21214 t

BASE COUNT 21203 a 14890 c 15769 g 21214 t

ORIGIN

Query Match 74.4%; Score 18.6; DB 4; Length 73076;  
 Best Local Similarity 84.0%; Pred. No. 2e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtgcagcgttg 25  
 ||| ||||| ||||| ||| |||

Db 17969 TCGACAGGTGCAGGTGCAGGTG 17993

RESULT 22  
 AC026904

LOCUS  
 DEFINITION Homo sapiens chromosome 6 clone RP11-567J20 map 6, WORKING DRAFT  
 SEQUENCE, 2 ordered pieces.

ACCESSION AC026904  
 VERSION AC026904.3 GI:12545335  
 KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
 SOURCE human.

ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 175455)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens chromosome 6, clone RP11-567J20  
 Unpublished

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS

2 (bases 1 to 175455)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
 Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,  
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Glade,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
 Melrim,J., Menues,L., Mihova,I., Miranda,C., Mienga,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neill,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,  
 Vassiliev,H., Videl,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission  
 Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jan 26, 2001 this sequence version replaced gi:7770479.  
 All repeats were identified using RepeatMasker:  
 Smit,A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information

Center project name: L7093  
 Center clone name: 567\_J\_20  
 ----- Summary Statistics  
 Sequencing vector: M13; M7815; 44% of reads  
 Sequencing vector: Plasmid; n/a; 56% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 174789 bases at least Q40  
 Consensus quality: 175132 bases at least Q30  
 Consensus quality: 175296 bases at least Q20  
 Insert size: 176000; agarose-fp  
 Insert size: 175355; sum-of-contents  
 Quality coverage: 9.4 in Q20 bases; agarose-fp  
 Quality coverage: 9.4 in Q20 b.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 167323: contig of 167323 bp in length  
 \* 167324 167423: gap of 100 bp  
 \* 167424 175455: contig of 8032 bp in length.

FEATURES  
 source  
 Location/Qualifiers  
 1. .175455  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /map="6"  
 /clone="RP11-567J20"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 1. .167323  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:left"  
 167424..175455  
 /note="assembly\_fragment  
 clone\_end:T7  
 vector\_side:right"  
 51313 a 39056 c 37387 g 47599 t 100 others

BASE COUNT 51313 a 39056 c 37387 g 47599 t 100 others

ORIGIN

Query Match 74.4%; Score 18.6; DB 70; Length 175455;  
 Best Local Similarity 84.0%; Pred. No. 1.7e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtgcagcgttg 25  
 ||| ||||| ||||| ||| |||

Db 137375 TCGTCTTGTGCAGGACACAGGTTG 137399

RESULT 23  
 AE003582/c

LOCUS  
 DEFINITION Drosophila melanogaster genomic scaffold 14200001386046 section 9  
 of 16, complete sequence.

ACCESSION AE003582  
 VERSION AE003582.2  
 KEYWORDS HTG.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 295312)  
 Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,  
 Ananides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,  
 George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,  
 Sutton,G.G., Wortman,J.R., Vandal,M.D., Zhang,Q., Chen,L.X.,

Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burris, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Foster, C., Gabriellian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattel, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Paclet, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.

The genome sequence of *Drosophila melanogaster*  
Science 287 (5461), 2185-2195 (2000)

2 (bases 1 to 295312)  
Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.  
Direct Submission  
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA  
On Oct 9, 2000 this sequence version replaced gi:7295900.  
Location/Qualifiers  
1. 295312  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/chromosome="2L"  
complement(join(16713..17645,17816..>18127))  
/gene="CG3515"  
/product="CT11855"  
/db\_xref="FLYBASE:FBan0003515"  
/db\_xref="FLYBASE:FBgn0031431"  
complement(<16713..>18127)  
/gene="CG3515"  
/map="22F3-23A1"  
/db\_xref="FLYBASE:FBan0003515"  
/db\_xref="FLYBASE:FBgn0031431"  
complement(join(16713..17645,17816..>18043))  
/gene="CG3515"  
/note="CG3515 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0003515"  
/db\_xref="FLYBASE:FBgn0031431"  
/protein\_id="AAF51244.1"  
/db\_xref="GI:7295945"  
/translation="MRQFRVAKVPETKHTLAMEFSSDRLRLVCDHSAISFSTR QVQVKPMHLYPDVACVQDTRVHSTKSHDYTRICLDNLRHCVRFVGHAKTI NRLAQPGGNNVDSISAGRDQVYMDIRVTKTRIKLRPLCPDPELFLATSSG TEGIOIRDVMMGSEPRKRVYQVNAKANWLOFAPNGKSLI1STDHSCFVSVAID GTYQSQFTGYSNKRLPLDATTYPDSQFILSGADKGRHWRADGFPVAVLKGNNMG PVKCMRPNRATWFSVDSRIWFMWPMWAGVVDWVKRLKPKETISFTDVTQVEAFAK PRTIPMPKPKIETVWKIKRNRKACTGDEIVDLTNALHNSDMEDETAED"

mRNA

complement(join(28221..28381,28449..28712,28820..28943,29016..29435,29487..29838,29893..30164,30434..30564,30620..>30841))  
/gene="Cyp309a1"  
/product="Ct28097"  
/db\_xref="FLYBASE:FBan0009964"  
/db\_xref="FLYBASE:FBgn0031432"  
complement(<28221..>30841)  
/gene="Cyp309a1"  
/note="CG9964"  
/map="22F4-23A1"  
/db\_xref="FLYBASE:FBan0009964"  
/db\_xref="FLYBASE:FBgn0031432"  
complement(join(28221..28381,28449..28712,28820..28943,29016..29435,29487..29838,29893..30092))  
/gene="Cyp309a1"  
/note="Cyp309a1 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0009964"  
/db\_xref="FLYBASE:FBgn0031432"  
/protein\_id="AAF51243.2"  
/db\_xref="GI:1072730"

gene

CDS

mRNA

/translation="MFTLVGLCLTIVHFAVYVLYTWYHKYWDKRGVWTAEPDLTIL GSPGILINKSRLLDQDVNYKDKYRTVGTFTROPQLLVLDPALAHEILVDKF SHFRDITTSFVGHNPDDKYVAGSPFFSAGDKWKRLENYVGLDTPSLKWAYSIWEQ SKRKLYEIERARREGQDIETRLDAYREFTANAMADFTWGDAGLSGKVGIEGDFQK TSTDIAIAFSFMRFNKTLVAIFVRKLFMSKREFTKATDEFFRLTQDAVNLROGGSG ERTDYLSHLLOQOGRNSIHDSVGHALTVHLGDGFETSGAVLYHMLYSLSPHREOEK LRSELEALASEGQISYDQINNLPLYLDCFNESRLTTPIGCFEFTKTPQINLGGD KTLDEPGVTVMVPAITQTHNDIYPEASEFSPORFENGASVLTIRKGCFLPQDGPRI CLGMRVGQLSVKTAIVHLSNYQVEQMKKVPFGADSGMGLNGDVELKTKLQK" 38658..>38947))  
/gene="CG18559"  
/product="CT41535"  
/db\_xref="FLYBASE:FBan0018559"  
/db\_xref="FLYBASE:FBgn0031433"  
/evidence=not\_experimental  
complement(<37517..>38947)  
/gene="CG18559"  
/map="23A1-23A1"  
/db\_xref="FLYBASE:FBan0018559"  
/db\_xref="FLYBASE:FBgn0031433"  
/evidence=not\_experimental  
complement(join(37517..37711,37769..38191,38253..38604,38658..>38947))  
/gene="CG18559"

gene

CDS

complement(join(37517..37711,37769..38191,38253..38604,38658..>38947))  
/gene="CG18559"  
/note="CG18559 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0018559"  
/db\_xref="FLYBASE:FBgn0031433"  
/evidence=not\_experimental  
/protein\_id="AAF51242.1"  
/db\_xref="GI:7295943"  
/translation="MMPERFSDFRPIDQVWYTRAKGACSNTRKGNMVLASALILHL LVLPYLYLTHHHKRWKRGVLTARPLTLGLTYPGLLTRKSNLVDFQKIYDKYK HRAVCVVRTRQPLVLDPELAHEVLSNFRCYKDSQSLRSHAKWDTARLNPFWA SQSWRRRTDAQIGISGRQAYNIWEGQMLTETMTQQAENKNIETRLDCLCFR YTAHVMAFDIWDIDAGTLTRPEQPNKVOEMASKWTSYAFMLTLFNMATIVAPCSRL LRFKPYKETDEFFSNLTKESELKRGADSTRDYLSHLQLRQKQATHDDLVGHA LTVMLDGYDTSCTALLHALYLAENPAVQVKLRVLEILSCMASEKSLDFELSSIQYLE QTVYKHTYRLCOFHIIYIVKLLFYL"

mRNA

complement(join(39362..40140,40214..40753,40814..40989))  
/gene="Rad1"  
/product="CT10890"  
/db\_xref="FLYBASE:FBan0003240"  
/db\_xref="FLYBASE:FBgn0026778"  
complement(<39362..>40989)  
/gene="Rad1"  
/note="CG3240"  
/map="23A1-23A1"  
/db\_xref="FLYBASE:FBan0003240"  
/db\_xref="FLYBASE:FBgn0026778"

gene



```

repeat_region 13252. .13483
/note="MIR repeat: matches 7. .243 of consensus"
repeat_region 13883. .14071
/note="MIR repeat: matches 64. .261 of consensus"
repeat_region 14118. .14425
/note="Alusx repeat: matches 1. .308 of consensus"
repeat_region 14451. .14547
/note="L2 repeat: matches 2614. .2710 of consensus"
repeat_region 14638. .14848
/note="MIR repeat: matches 22. .233 of consensus"
repeat_region 14863. .15592
/note="L2 repeat: matches 1917. .2732 of consensus"
repeat_region 16139. .16203
/note="L2 repeat: matches 2577. .2664 of consensus"
repeat_region 16736. .16776
/note="L2 repeat: matches 2657. .2697 of consensus"
repeat_region 16937. .16980
/note="L1 copies 4 mer acat 77% conserved"
repeat_region 18563. .18688
/note="Alusg/x repeat: matches 186. .310 of consensus"
repeat_region 19350. .19611
/note="Alusg repeat: matches 41. .313 of consensus"
repeat_region 19917. .20052
/note="L2 repeat: matches 2617. .2742 of consensus"
repeat_region 20053. .20248
/note="Alusg/x repeat: matches 91. .285 of consensus"
repeat_region 20249. .20291
/note="L2 repeat: matches 2578. .2617 of consensus"
repeat_region 20292. .20582
/note="AlusC repeat: matches 1. .291 of consensus"
repeat_region 20583. .20942
/note="L2 repeat: matches 2160. .2578 of consensus"
repeat_region 22096. .22252
/note="MIR repeat: matches 24. .191 of consensus"
repeat_region 22924. .23084
/note="MIR repeat: matches 32. .201 of consensus"
repeat_region 23273. .23453
/note="L1ME repeat: matches 5603. .5802 of consensus"
repeat_region 23747. .23956
/note="MER20 repeat: matches 6. .218 of consensus"
repeat_region 24553. .24771
/note="MER46A repeat: matches 1. .236 of consensus"
repeat_region 25172. .25464
/note="AluIo repeat: matches 1. .302 of consensus"
misc_feature complement(25399. .25912)
/note="match: GSS: Em:AQ381838"
misc_feature complement(25464. .25902)
/note="match: GSS: Em:AQ121579"
repeat_region 25491. .25613
/note="MLR11 repeat: matches 287. .408 of consensus"
repeat_region 25759. .25858
/note="MLR11 repeat: matches 1. .90 of consensus"
repeat_region 26095. .26289
/note="MIR repeat: matches 3. .233 of consensus"
repeat_region 27202. .27364
/note="MIR repeat: matches 79. .252 of consensus"
repeat_region 27684. .28115
/note="MLR11 repeat: matches 36. .472 of consensus"
repeat_region 28295. .28518
/note="MIR repeat: matches 29. .250 of consensus"
repeat_region 29676. .29957
/note="L2 repeat: matches 2252. .2520 of consensus"
repeat_region 30268. .30557
/note="Alusx repeat: matches 1. .293 of consensus"
repeat_region 30558. .30693
/note="L17 copies 8 mer tatatata 84% conserved"
repeat_region 30560. .30695
/note="L34 copies 4 mer tata 85% conserved"
repeat_region 31523. .31813
/note="Alusg repeat: matches 1. .291 of consensus"
repeat_region 31917. .31960
/note="L22 copies 2 mer ac 93% conserved"
repeat_region 31918. .31961
/note="L1 copies 4 mer caca 95% conserved"
31920. .31959
/note="5 copies 8 mer cacacaca 97% conserved"
32433. .32745
/note="L1ME2 repeat: matches 5860. .6164 of consensus"
32820. .33044
/note="MER1B repeat: matches 125. .337 of consensus"
33045. .33306
/note="L1MB7 repeat: matches 5628. .5877 of consensus"
33333. .33788
/note="L1MB7 repeat: matches 5729. .6173 of consensus"
34018. .34081
/note="L2 repeat: matches 2 mer ta 75% conserved"
34024. .34079
/note="7 copies 8 mer tatatata 78% conserved"
35097. .35280
/note="MER69 repeat: matches 74. .264 of consensus"
36056. .36229
/note="L2 repeat: matches 2358. .2534 of consensus"
36230. .36449
/note="MER58A repeat: matches 1. .220 of consensus"
36450. .36583
/note="L2 repeat: matches 2533. .2666 of consensus"
37373. .37723
/note="MER1B repeat: matches 1. .364 of consensus"
38181. .38399
/note="L2 repeat: matches 2457. .2656 of consensus"
38402. .38745
/note="MER1B repeat: matches 1. .337 of consensus"
38884. .39041
/note="MER45 repeat: matches 5. .178 of consensus"

Query Match 72.8%; Score 18.2; DB 93; Length 72889;
Best Local Similarity 87.0%; Pred. No. 3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 gtcaggtgcaggtcacgcagcttg 25
|||||
Db 38249 GTGAGGTGCAGGTAACGACGTGG 38271

RESULT 25
AC003096 72919 bp DNA PLN 05-APR-2000
LOCUS Arabidopsis thaliana chromosome II section 191 of 255 of the
DEFINITION complete sequence. Sequence from clones T31E10, T29F13.
ACCESSION AC003096 AE002093
VERSION AC003096.2 GI:6598389
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 72919)
AUTHORS Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,
Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
Felchlyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L.,
Tallon,B.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
Venter,J.C.
Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana
JOURNAL Nature 402 (6763), 761-768 (1999)
MEDLINE 20083487
PUBMED 10617197
REFERENCE 2 (bases 1 to 72919)
AUTHORS Lin,X.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA

```

COMMENT

On Dec 17, 1999 this sequence version replaced gi:3132469.  
The sequence and annotation of chromosome 2 were merged from those  
of the individual clones on this chromosome after removing  
overlaps. For detailed information, please see the TIGR web site  
(<http://www.tigr.org/tdb/at/at.html>).

Genes were identified by a combination of three methods: Gene  
prediction programs including GRAIL  
(<ftp://arthur.epm.ornl.gov/pub/xgrail>), GeneFinder (Phil Green,  
University of Washington), Genscan (Chris Burge,  
<http://gnomic.stanford.edu/GENSCAN.html>), and NetPlantGene  
(<http://www.cbs.dtu.dk/services/NetPlantGene/>), searches of the  
complete sequence against a peptide database and plant EST  
databases at TIGR, and manual curations based on those analyses.  
Annotated genes are named to indicate the level of evidence for  
their annotation. Genes with similarity to other proteins are named  
after the database hits. Genes without significant peptide  
similarity but with EST similarity are named as 'unknown' proteins.  
Genes without protein or EST similarity, that are predicted by two  
or more gene prediction programs over most of their length are  
annotated as 'hypothetical' proteins. Genes encoding tRNAs are  
predicted by tRNAscan-SE (Sean Eddy,  
<http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were  
identified by RepeatMasker (Arian Smit,  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are  
numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones  
F6P23, F536, T17A5, and T31L6, the ESSA group for sequencing clone  
F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards  
and Satoshi Tabata for helpful assistance. In addition, we would  
like to thank the TIGR Bioinformatics Department, especially Lixin  
Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy  
Peterson, Michael Holmes, and Delwood Richardson for software and  
database support.

This work was supported by the National Science Foundation,  
Department of Energy and the US Department of Agriculture.

Address all correspondence to: [at@tigr.org](mailto:at@tigr.org).

FEATURES

source	Location/Qualifiers
	1..72919
	/organism="Arabidopsis thaliana"
	/cultivar="Columbia"
	/db_xref="taxon:3702"
	/chromosome="II"
misc_feature	complement(<1..5635)
repeat_region	/note="Sequence from clone T31E10"
repeat_region	complement(234..349)
repeat_region	/rpt_family="CAAAA)n"
repeat_region	464..491
repeat_region	/rpt_family="POLY_A"
repeat_region	1667..1709
repeat_region	/rpt_family="(TA)n"
repeat_region	join(5636..6518,6612..6790,6877..6916,7039..7195, 7361..7475,7554..7634,8023..8047,8274..8330,8425..8490, 8771..8845,9250..9357,9777..9911,10048..10154, 10232..10325,10410..10577,10752..10892,10993..11193, 11414..11512,11613..11765,11888..12285,12468..12648, 12857..13180,13384..14262,14338..14583,14677..15091, 15212..15488)
repeat_region	/gene="At2g34660"
repeat_region	5636..15488
repeat_region	/gene="At2g34660"
repeat_region	/note="T29F13.13; identical to GB:AF014960; transports glutathione conjugates"
repeat_region	complement(5636..72919)
repeat_region	/note="Sequence from clone T29F13"
repeat_region	join(6189..6518,6612..6790,6877..6916,7039..7195, 7361..7475,7554..7634,8023..8047,8274..8330,8425..8490, 8771..8845,9250..9357,9777..9911,10048..10154, 10232..10325,10410..10577,10752..10892,10993..11193, 11414..11512,11613..11765,11888..12285,12468..12648,

12857..13180,13384..14262,14338..14583,14677..15091, 15212..15309) /gene="At2g34660" /codon_start=1 /product="ABC transporter (ALMRP2)" /protein_id="AAC16268.1" /db_xref="GI:3132479" /translation="MGFEIEWYCKPVPNCVWTKOVANAFGATPCATDSFVLGTSOL VLIVLCYRIWLAKDKHVERCLSRSLYFIALLAAATAPLEFLRMGTSVLDFO GPGLPPEAFGLGKAFKAWGAVMILMETKIYIRLWYVFAVIALVGLDMVLINL VLSKEYYSYVLYLTVSEGAQVLFGLIFLHLPNLDYPGYVRSYVDDYEEY ISDQOICEKHANIFDKIFFSMNPMTLGSKRPITKEDVYLDTDOTETLFTSFO HWDKLEQKOPHLLBALANSGLGRWGGFWKIGNDSQFVGLPILNOLLKSMQEDA PAMWGIYAFISFVGVFVCLCAQFQNMVGVYRSLAALAAVPRKSLRLNREGR KFTGKITNLTDAESLQOICSLHMTSAPRIIILILLYQOGLVSLGALLV LMPEPTVITISKMKLTKEGLQRTDKRIGLMNEVLAAMDTVKYAMENSFQKQVTVR DBLSWFRKSQLGALNMTILNSIPVLVTIVSGVFTLLGPTLTPARATLSLFAVL RFLPLNITTOVANNVSLKRLVLAETEEIRLLPNPILPEGPAIIRINGYFSWD SKGDRPLTSLNIDVPLGSLVAVVGSGKTSLSAILGELPATSDATVTLRGSVAY VPOVSIFATVVDNLFGSPFDREYRAIDVTSKHLDELPGDGLTEIGRGNVI SGQQRVSMARAVISNDVIIPDDPLSADAHVGOQVEKIKRELQKTRVTVNQ LHFSQVDRITVHEGTVEEGTYEELSSNGPLFORLMENAGKVEEYSENGEAEADQ TAPQVANGTNGLODGGDKKSGKNGKSVLIKQERETGVVSWRLKRYODA LGCAVWMLLCYVLTVEFRTSTWLSSEWTDAGTKPGHPLFYLIYALLSFGVL VLTNSYWLIMSSLYAAKLIHDMHLSILRAPMSPFTNPLGRIINRAKLDGIDRT VAYVNMFCQVSQLSTVYVLTIGVSTLSIWMPLLVLYGAYLYQNTAREVRMD TSRSPYIAQGEALNSTIRAKYADRMADNGRSMONRTFLVNMGNARLGIIR LETGLMTASTFAVNGRAENQOQAFSTMGILLSYLNITSLTGLVLRSLAE NSLNARVGVNIEIPPEAPVTENNRPPGPPSGSIRKEDVLRYPQLPPLHGV SFTHPTDKVIGIRPGACKSLLNALRIVEVEKGRILIDDCDCKGFCGLMDLRKVLG IIPQVPLSGTVRENDLPDGHENDLWESLRAHDKTIRNPLGLDAEYSEAGEN FSVGQQLLSALLRRSKILYLDATRAVDYRTALQKTIREPEKSCYMLTIAHR LNTLIDCKILVLDGSRVQFSPFNLLSGSSFKMSVFGAANAEYVTLTQVLTDRK RAKDDSHHGOGRKWLASRAAAQAFALAASTLSSHNDLSLETDEDDSSVRLDNRK VATRSVLECKHDKHKAESLEENISREGWLSLYRWVEGLAVMSLRARNMQQPDYN FEGTDFDQNVEM"	misc_feature complement(13668..14027) /note="molecular marker ve016" complement(join(<15663..16061,16251..16560,16668..16891, 16973..17154,17274..>18153)) /gene="At2g34670" complement(<15663..>18153) /gene="At2g34670" /note="T29F13.12" complement(join(15663..16061,16251..16560,16668..16891, 16973..17154,17274..>18153)) /gene="At2g34670" /note="hypothetical protein" /codon_start=1 /protein_id="AAC16267.1" /db_xref="GI:3132478" /translation="MGCAASRIDNEKVLVCRQKRLMKLLGFRGEFADQAQYLA LRNTGTVLRQFTESETLENTSYGLSLPLPPPTLPSPPPPSPPPPSPPPPSP HDLADEEGENDGNDGAAPPPLPNSWINPWFESLELHSHNGPNVYOVLEK KQOIOAEEDWAETKQFEEDEOQAGGCLDLVHQIEAVSGCNMKPRRLKFK LGEVMDGNSMFTSCSKDKLEKTHVTDCTIRRTLEGLIRELDYDFLKASCCEKIAIV DINSRTOVDFRTRRRKSSKVSFSAWSWSKSQLQGLDQKATSCCEKIAIV HCTLEKLYTAKKLYQLYRKEAHEKRSALQKQDGETYDLKSKMKARLSLE LETEIQRLEDSTTRSCLLNINDELYPQVLTVALTSGLAQMKMTMLKHQVHIHSQ LNLHPYPSIDLSSEYKQAVNEITEVTCWNSFKLVNSQVYVTKTCTVQLTDRK LSNEDNQRSLPVAARKLCKEQLVNLKCNKLERLEKELISLAETERRLEGILA MEEEVSYSLGSKHPLSLKQAKIARLRKVDIEKTYLINSVEVSKRMTLDNLKSLP NVFMULTANFANGFESVNGTGTVDTSQHSDESOP"
	repeat_region complement(17583..17677) /rpt_family="(GAA)n" /rpt_family="(GGA)n" 18888..18927 /rpt_family="POLY_A" complement(join(19728..20045,20129..20260,20346..20435, 20465..20523,20774..20905,21067..21222,21309..21490, 21666..21849,21949..22022,22320..22524,22781..22902, 22992..23187,23344..23463,23557..23630,23712..23796,

```

23837. 24080,24191. 24309,24596. 24772,25433. 25550,
25630. 25814,26314. 26414,26518. 26653,27022. 27197,
27322. 27467,27751. 27983,28270. 28371,28832. 28931,
29231. 29296,29387. 29449,29582. 29647,29732. 29803,
29894. 29959,30240. 30311,30410. >31222))
/gene="At2g34680"
complement(19728. >31222)
/gene="At2g34680"
/note="T29F13.11"
complement(join(19869. 20045,20129. 20260,20346. 20435,
20465. 20523,20774. 20905,21067. 21222,21309. 21490,
21666. 21849,21949. 22022,22320. 22524,22781. 22902,
22992. 23187,23344. 23463,23557. 23630,23712. 23796,
23837. 24080,24191. 24309,24596. 24772,25433. 25550,
25630. 25814,26314. 26414,26518. 26653,27022. 27197,
27322. 27467,27751. 27983,28270. 28371,28832. 28931,
29231. 29296,29387. 29449,29582. 29647,29732. 29803,
29894. 29959,30240. 30311,30410. 31222))
/gene="At2g34680"
/note="unknown protein"
/codon_start=1
/protein_id="AAC16266.1"
/db_xref="GI:3132477"
/translation="MEEVAARVEETVETNVDAVKEDNATIANESRSPESVAVSVVS
Query Match 72.8%; Score 18.2; DB 12; Length 72919;
Best Local Similarity 87.0%; Pred. No. 3e+02; 3; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 3;
QY 1 tcgtcaggtgcaggtcagcagt 23
II III IIIII IIIIIIIII
Db 55600 TCCTCAAGTGCATGTCAGCACGT 55622

RESULT 26
LOCUS ATT20010 83122 bp DNA PLN 13-APR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 3, BAC clone T20010.
ACCESSION AL163816
VERSION AL163816.1 GI:7573419
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 83122)
Obermaier,B., Ottenwaelder,B., Duchemin,D., Zeitler,K., Mewes,H.W.,
Rudd,S., Lemcke,K., Mayer,K.F.X., Quettier,F. and Salanoubat,M.
Unpublished
2 (bases 1 to 83122)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (13-APR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
Coordinator: Marcel Salanoubat and Francis Quettier, Grouperment
d'Interet Public, Centre National de Sequencage - GENOSCOPE, 2 rue
Gaston Cremieux, BP191, 91006 Evry Cedex, France;
http://www.genoscope.cns.fr
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
Location/Qualifiers
1. 83122
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
547. 3272
/gene="T20010_10"
complement(join(547. 627,715. 798,874. 921,1019. 1076,
1175. 1307,1385. 1457,1558. 1709,1847. 1912,2119. 2172,
2319. 2370,2485. 2588,2679. 2736,2811. 2891,2971. 3027,
3114. 3272))
/gene="T20010_10"
complement(547. 627)
/gene="T20010_10"
/number=1
complement(join(547. 627,715. 798,874. 921,1019. 1076,
1175. 1307,1385. 1457,1558. 1709,1847. 1912,2119. 2172,
2319. 2370,2485. 2588,2679. 2736,2811. 2891,2971. 3027,
3114. 3272))
/gene="T20010_10"
/note="strong similarity to translation releasing factor
RF-1, Synechocystis sp., PIR:S76914
Contains Prokaryotic-type class I peptide chain release
factors signature AA280-296"
/codon_start=1
/product="translation releasing factor RF-1-like protein"
/protein_id="CAB87736.1"
/db_xref="GI:7573420"
/translation="MNSMTTPTARAFHLHSSPTLRSRKSQSGVCLVSPRSR
RTRRLVCAEPYLLIRKMSVEKTKELSVKLADPDVSNQSEYOKLAQMSLEDRVST
VFRFKDCQKLESKVLAKEAGDDEDMAGMIGSEINSLTKEIELEKQLKVLPLPSD
PLDARNILLEVAGTGGDEAAIWTGDLGVADYSEVRAQLLEVFNGLMGVTKCVME
IKGNRYISKLYESGVHRVQVPTQGTQVHTSTATVAIMPEADEVEVVDPKDIEL
TSRSGGAGQNVNKVETAIIDLFRKPSGIRIFCTEERTQIRNKARAFOLLAKLYEIK
VREOQEKIRNERKSOVGTGARSEKIRTYNKDSRVTDHRLKMFALTFLDGALEDAV
QCAALEQKELMEELSESVAAASATSG"
complement(628. 714)
/number=1
complement(715. 798)
/gene="T20010_10"
/number=2
complement(799. 873)
/number=2
complement(874. 921)
/gene="T20010_10"
/number=3
complement(922. 1018)
/number=3
complement(1019. 1076)
/gene="T20010_10"
/number=4
complement(1077. 1174)
/number=4
complement(1175. 1307)
/gene="T20010_10"
/number=5
complement(1308. 1384)
/number=5
complement(1385. 1457)
/gene="T20010_10"
/number=6
complement(1458. 1557)
/number=6
complement(1558. 1709)
/gene="T20010_10"
/number=7
complement(1710. 1846)
/number=7
complement(1847. 1912)
/gene="T20010_10"
/number=8
complement(1913. 2118)
/number=8
complement(2119. 2172)
/gene="T20010_10"
/number=9
complement(2173. 2318)
/number=9
complement(2319. 2370)
/gene="T20010_10"
/number=10
complement(2371. 2484)

```



Center project name: L4035  
Center clone name: 21\_J\_18

\* NOTE: This record contains 91 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 894: contig of 894 bp in length  
895 994: gap of 100 bp  
995 1895: contig of 901 bp in length  
1896 1995: gap of 100 bp  
1996 2929: contig of 934 bp in length  
2930 3029: gap of 100 bp  
3030 3960: contig of 931 bp in length  
3961 4060: gap of 100 bp  
4061 5031: contig of 971 bp in length  
5032 5131: gap of 100 bp  
5132 6053: contig of 922 bp in length  
6054 6153: gap of 100 bp  
6154 7062: contig of 909 bp in length  
7063 7162: gap of 100 bp  
7163 8056: contig of 894 bp in length  
8057 8156: gap of 100 bp  
8157 9085: contig of 929 bp in length  
9086 9185: gap of 100 bp  
9186 10128: contig of 943 bp in length  
10129 10228: gap of 100 bp  
10229 11163: contig of 935 bp in length  
11164 11263: gap of 100 bp  
11264 12165: contig of 902 bp in length  
12166 12265: gap of 100 bp  
12266 13196: contig of 931 bp in length  
13197 13296: gap of 100 bp  
13297 14198: contig of 902 bp in length  
14199 14298: gap of 100 bp  
14299 15211: contig of 913 bp in length  
15212 15311: gap of 100 bp  
15312 16213: contig of 902 bp in length  
16214 16313: gap of 100 bp  
16314 17266: contig of 953 bp in length  
17267 17366: gap of 100 bp  
17367 18307: contig of 941 bp in length  
18308 18407: gap of 100 bp  
18408 19326: contig of 919 bp in length  
19327 19426: gap of 100 bp  
19427 20356: contig of 930 bp in length  
20357 20456: gap of 100 bp  
20457 21367: contig of 911 bp in length  
21368 21467: gap of 100 bp  
21468 22359: contig of 892 bp in length  
22360 22459: gap of 100 bp  
22460 23377: contig of 918 bp in length  
23378 23477: gap of 100 bp  
23478 24384: contig of 907 bp in length  
24385 24484: gap of 100 bp  
24485 25418: contig of 934 bp in length  
25419 25518: gap of 100 bp  
25519 26448: contig of 930 bp in length  
26449 26548: gap of 100 bp  
26549 27508: contig of 960 bp in length  
27509 27608: gap of 100 bp  
27609 28530: contig of 922 bp in length  
28531 28630: gap of 100 bp  
28631 29542: contig of 912 bp in length  
29543 29642: gap of 100 bp  
29643 30587: contig of 945 bp in length

30588 30687: gap of 100 bp  
30688 31615: contig of 928 bp in length  
31616 31715: gap of 100 bp  
31716 32650: contig of 935 bp in length  
32651 32750: gap of 100 bp  
32751 33680: contig of 930 bp in length  
33681 33780: gap of 100 bp  
33781 34707: contig of 927 bp in length  
34708 34807: gap of 100 bp  
34808 35781: contig of 974 bp in length  
35782 35881: gap of 100 bp  
35882 36784: contig of 903 bp in length  
36785 36884: gap of 100 bp  
36885 37813: contig of 929 bp in length  
37814 37913: gap of 100 bp  
37914 38865: contig of 952 bp in length  
38866 38965: gap of 100 bp  
38966 39920: contig of 955 bp in length  
39921 40020: gap of 100 bp  
40021 40937: contig of 917 bp in length  
40938 41037: gap of 100 bp  
41038 41980: contig of 943 bp in length  
41981 42080: gap of 100 bp  
42081 42997: contig of 917 bp in length  
42998 43097: gap of 100 bp  
43098 43997: contig of 900 bp in length  
43998 44097: gap of 100 bp  
44098 45007: contig of 910 bp in length  
45008 45107: gap of 100 bp  
45108 46068: contig of 961 bp in length  
46069 46168: gap of 100 bp  
46169 47090: contig of 922 bp in length  
47091 47190: gap of 100 bp  
47191 48084: contig of 894 bp in length  
48085 48184: gap of 100 bp  
48185 49107: contig of 923 bp in length  
49108 49207: gap of 100 bp  
49208 50167: contig of 960 bp in length  
50168 50267: gap of 100 bp  
50268 51211: contig of 944 bp in length  
51212 51311: gap of 100 bp  
51312 52258: contig of 947 bp in length  
52259 52358: gap of 100 bp  
52359 53260: contig of 902 bp in length  
53261 53360: gap of 100 bp  
53361 54277: contig of 917 bp in length  
54278 54377: gap of 100 bp  
54378 55306: contig of 929 bp in length  
55307 55406: gap of 100 bp  
55407 56360: contig of 954 bp in length  
56361 56460: gap of 100 bp  
56461 57356: contig of 896 bp in length  
57357 57456: gap of 100 bp  
57457 58417: contig of 961 bp in length  
58418 58517: gap of 100 bp  
58518 59425: contig of 908 bp in length  
59426 59525: gap of 100 bp  
59526 60442: contig of 917 bp in length  
60443 60542: gap of 100 bp  
60543 61447: contig of 905 bp in length  
61448 61547: gap of 100 bp  
61548 62497: contig of 950 bp in length  
62498 62597: gap of 100 bp  
62598 63505: contig of 908 bp in length  
63506 63605: gap of 100 bp  
63606 64553: contig of 948 bp in length  
64554 64653: gap of 100 bp  
64654 65605: contig of 952 bp in length  
65606 65705: gap of 100 bp  
65706 66663: contig of 958 bp in length  
66664 66763: gap of 100 bp  
66764 67658: contig of 895 bp in length  
67659 67758: gap of 100 bp





```

/standard_name="SHGC-82724"
/evidence=not_experimental
/db_xref="dbSTS:G48687"
complement(50213..50359)
/rpt_family="MER45"
/rpt_type=dispersed
50367..122949
/note="similar to Homo sapiens clone RP11-21J18 deposited
in GenBank Accession Number AC022591.2"
50505..50741
/rpt_family="MER1"
/rpt_type=dispersed
50767..50842
/rpt_family="MER1"
/rpt_type=dispersed
complement(51523..51873)
/rpt_family="THE1"
/rpt_type=dispersed
complement(56228..56300)
/rpt_family="MER1"
/rpt_type=dispersed
complement(56315..56511)
/rpt_family="MER1"
/rpt_type=dispersed
complement(57433..57713)
/rpt_family="Alu"
/rpt_type=dispersed
complement(58689..58969)
/rpt_family="Alu"
/rpt_type=dispersed
63803..64058
/rpt_family="Alu"
/rpt_type=dispersed
64450..64608
/rpt_family="MER46"
/rpt_type=dispersed
complement(64504..64688)
/rpt_family="MER46"
/rpt_type=dispersed
complement(65294..65499)
/rpt_family="MER20"
/rpt_type=dispersed
67033..67145
/rpt_family="MIR"
/rpt_type=dispersed
complement(68663..68943)
/rpt_family="Alu"
/rpt_type=dispersed
complement(68995..69186)
/rpt_family="Alu"
/rpt_type=dispersed
69649..69909
/rpt_family="Alu"
/rpt_type=dispersed
complement(70559..70669)
/rpt_family="Alu"
/rpt_type=dispersed
complement(74449..74549)
/rpt_family="MIR"
/rpt_type=dispersed
74819..75109
/rpt_family="Alu"
/rpt_type=dispersed
complement(75767..75983)
/rpt_family="MIR"
/rpt_type=dispersed
complement(77660..77936)
/rpt_family="Alu"
/rpt_type=dispersed
complement(79568..79670)
/rpt_family="Alu"
/rpt_type=dispersed
80832..80988

```

```

/rpt_family="Alu"
/rpt_type=dispersed
82365..82843
repeat_region
Query Match 72.8% Score 18.2; DB 89; Length 122951;
Best Local Similarity 87.0%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 3 qtcagggtcaggtcagcagcttg 25
||| ||||| ||||| ||||| |||||
Db 50997 GTGAGGTGACGGTAAGCACGTGG 50975

RESULT 29
AC004691/c
LOCUS AC004691 141990 bp DNA PRI 03-FEB-2000
DEFINITION Homo sapiens PAC clone RP4-740D2 from 7p14-p15, complete sequence.
ACCESSION AC004691
VERSION AC004691.1 GI:3135285
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Dempsey,S., Lacy,M., Holmes,A. and Nguyen,C.
The sequence of Homo sapiens PAC clone RP4-740D2
Unpublished
REFERENCE 2 (bases 1 to 141990)
Waterston,R.
Direct Submission
Submitted (16-MAY-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
3 (bases 1 to 141990)
Waterston,R.
Direct Submission
Submitted (03-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_DJ0740D02
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-4, prepared by Pieter de Jong and coworkers at Roswell Park Cancer Institute, using the method described by Ioannou et al., Nature Genetics

6:84-9 (1994). The library is from one male donor. For further details, see <http://pacpac.med.buffalo.edu/> (http://www.genomesystems.com).  
 VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:  
 The clone sequenced to the right is RP5-877J2. Actual start of this clone is at base position 1 of RP4-740D2; actual end is at 141990 of RP4-740D2.

This clone contains SPS SWSS1731 (NID:gl113195).

#### FEATURES

##### Location/Qualifiers

```

1..141990
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7p14-pl5"
/clone="RP4-740D2"
/clone_lib="RPC1-4"
751..915
/rpt_family="MIR"
1675..1765
/rpt_family="MIR"
1768..1957
/rpt_family="L2"
2057..3427
/rpt_family="L2"
3453..4205
/rpt_family="MER21_g"
4218..4956
/rpt_family="L2"
5208..5290
/rpt_family="L2"
5404..5567
/rpt_family="Alu"
10107..10264
/rpt_family="L2"
10631..10678
/rpt_family="MaLR"
10679..10726
/rpt_family="Retroviral"
10719..10804
/rpt_family="(CATA)n"
10815..11116
/rpt_family="Alu"
11118..11424
/rpt_family="Retroviral"
11446..11824
/rpt_family="MaLR"
11848..11984
/rpt_family="p2_type"
11986..12048
/rpt_family="MaLR"
14726..15039
/rpt_family="Alu"
16156..17566
/rpt_family="MER1_type"
18844..19139
/rpt_family="Alu"
19186..19256
/rpt_family="MIR"
20185..20389
/rpt_family="MER1_type"
20582..21615
/rpt_family="L1"
21811..22062
/rpt_family="Alu"
22336..22400
/rpt_family="MIR"
23870..23933
/rpt_family="MIR"
23985..24487
/rpt_family="MaLR"

```

```

repeat_region 25373..25574
/rpt_family="MER1_type?"
repeat_region 25583..26294
/rpt_family="L2"
repeat_region 26438..26697
/rpt_family="MER2_type"
26712..26847
/rpt_family="L2"
26893..27192
/rpt_family="Alu"
27979..28151
/rpt_family="MER1_type"
28152..28414
/rpt_family="Retroviral"
31181..31300
/rpt_family="(CCAA)n"
33019..33151
/rpt_family="MIR"
34032..34332
/rpt_family="L2"
35060..35102
/rpt_family="L2"
35922..36389
/rpt_family="Retroviral"
36390..37203
/rpt_family="Retroviral"
37283..37402
/rpt_family="(GA)n"
37404..37951
/rpt_family="Retroviral"
38016..38088
/rpt_family="MER4-group"
38095..38178
/rpt_family="Retroviral"
38291..38589
/rpt_family="Alu"
38728..39167
/rpt_family="MER4-group"
39298..39583
/rpt_family="Retroviral"
39670..39913
/rpt_family="Retroviral"
39915..40376
/rpt_family="Retroviral"
40534..40815
/rpt_family="L1"
42486..42671
/rpt_family="Achobo"
42860..43006
/rpt_family="MIR"
46335..46765
/rpt_family="L1"
46795..47028
/rpt_family="MIR"
47156..47719
/rpt_family="L1"
47779..48269
/rpt_family="L1"
48295..48595
/rpt_family="Alu"
48778..49324
/rpt_family="L1"
49647..50115
/rpt_family="L1"
50261..50378
/rpt_family="L1"
50415..50545
/rpt_family="L1"
50571..50865

```

Query Match 72.8%; Score 18.2; DB 85; Length 141990;  
 Best Local Similarity 87.0%; Pred. No. 2.7e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY      3 gtcagggtcagggtcagcagcttg 25
||||| ||||| ||||| ||||| |||||
Db 103543 GTCAGGTGCATGCTGCACATTG 103521

RESULT 30
AC007044
LOCUS   AC007044      180257 bp      DNA      HTG      10-SEP-2000
DEFINITION Homo sapiens chromosome UNK clone CTA-435J10, WORKING DRAFT
ACCESSION AC007044
VERSION   AC007044.3 GI:10048128
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
           Waterston,R.H.
           The sequence of Homo sapiens clone
           Unpublished
           2 (bases 1 to 180257)
           Waterston,R.H.
           Direct Submission
           Submitted (06-MAR-1999) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
           On Sep 10, 2000 this sequence version replaced gi:9838354.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Center project name: H_RG435J10
----- Project Information -----
----- Summary Statistics -----
Sequencing vector: M13; %
Sequencing vector: plasmid; %
Chemistry: Dye-terminator Big Dye; % of reads
Chemistry: Dye-terminator Big Dye; % of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 180641 bases at least Q40
Consensus quality: 181075 bases at least Q30
Consensus quality: 181328 bases at least Q20
Insert size: 159000; agarose-fp
Insert size: 181700; sum-of-contigs
Quality coverage: in Q20 bases; agarose-fp
Quality coverage: in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 180257: contig of 180257 bp in length.
*
FEATURES             source
    source            Location/Qualifiers
    ..180257
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="UNK"
    /clone="CTA-435J10"
    misc_feature      1. 180257
    /note="assembly_name:Contig36
    clone_end:SP6
    vector_side:right"
    BASE COUNT        46943 a 45989 c 45264 g 42061 t
    ORIGIN

```

---

```

Query Match      72.8%; Score 18.2; DB 60; Length 180257;
Best Local Similarity 87.0%; Pred. No. 2.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 gtcagggtcagggtcagcagcttg 25
||||| ||||| ||||| ||||| |||||
Db 178754 GTCAGGTGCATGCTGCACATTG 178776

RESULT 31
AC064835
LOCUS   AC064835      202828 bp      DNA      HTG      07-JUL-2000
DEFINITION Homo sapiens chromosome 16 clone RP11-673P17, WORKING DRAFT
ACCESSION AC064835
VERSION   AC064835.3 GI:7839924
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
           Waterston,R.H.
           The sequence of Homo sapiens clone
           Unpublished
           2 (bases 1 to 202828)
           Waterston,R.H.
           Direct Submission
           Submitted (22-APR-2000) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
           On May 16, 2000 this sequence version replaced gi:7712280.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Center project name: H_NH0673P17
----- Project Information -----
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 195107 bases at least Q40
Consensus quality: 197538 bases at least Q30
Consensus quality: 198979 bases at least Q20
Insert size: 199000; agarose-fp
Insert size: 201728; sum-of-contigs
Quality coverage: 5.33 in Q20 bases; agarose-fp
Quality coverage: 5.29 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1109: contig of 1109 bp in length
* 1110 1209: gap of unknown length
* 1210 5935: contig of 4726 bp in length
* 5936 6035: gap of unknown length
* 6036 13020: contig of 6985 bp in length
* 13021 13120: gap of unknown length
* 13121 21028: contig of 7908 bp in length
* 21029 21128: gap of unknown length
* 21129 27414: contig of 6285 bp in length
* 27414 34551: gap of unknown length
* 34551 34651: contig of 7038 bp in length
* 34651: gap of unknown length

```





FEATURES	source	California, Berkeley, 111 Koshland Hall, Berkeley, CA 94720, USA
Location/Qualifiers	1. .694	
/organism="metal-contaminated soil clone K20-78"		
/db_xref="taxon:95834"		
/clone="K20-78"		
<l. .>694		
/product="16S ribosomal RNA"		
179 a 150 c 237 g 128 t		
BASE COUNT		
ORIGIN		
Query Match	70.4%;	Score 17.6; DB 1; Length 694;
Best Local Similarity	83.3%;	Prod. No. 1.3e+03;
Matches 20; Conservative	0;	Mismatches 4; Indels 0; Gaps 0;
QY	2	cgtcagggtcaggtcagcacttg 25
Db	679	CCTCAGATGCAGCCAGCAGCTCG 656
RESULT 34		
AF269004/c		
LOCUS		
DEFINITION	AF269004	1412 bp DNA BCT 16-FEB-2001
		Uncultured bacterium SBR1108 16S ribosomal RNA gene, partial
		sequence.
ACCESSION	AF269004	X84534
VERSION	AF269004.1	GI:12751010
KEYWORDS		uncultured bacterium SBR1108.
SOURCE		uncultured bacterium SBR1108
ORGANISM		Bacteria; Green non-sulfur bacteria; environmental samples.
REFERENCE		1 (bases 1 to 1412)
AUTHORS		Bond,P.L., Hugenholtz,P., Keller,J. and Blackall,L.L.
TITLE		Bacterial community structures of phosphate-removing and non-phosphate-removing activated sludges from sequencing batch reactors
JOURNAL		Appl. Environ. Microbiol. 61 (5), 1910-1916 (1995)
MEDLINE		95374005
PUBMED		7544094
REFERENCE		2 (bases 1 to 1412)
AUTHORS		Hugenholtz,P., Tyson,G.W., Webb,R.I., Wagner,A.M. and Blackall,L.L.
TITLE		Investigation of candidate division TW7, a recently recognized major lineage of the domain bacteria with no known pure-culture representatives
JOURNAL		Appl. Environ. Microbiol. 67 (1), 411-419 (2001)
MEDLINE		20575237
PUBMED		11133473
REFERENCE		3 (bases 1 to 1412)
AUTHORS		Bond,P.L.
TITLE		Direct Submission
JOURNAL		Submitted (02-FEB-1995) Department of Chemical Engineering, University of Queensland, Brisbane, Queensland 4072, Australia
REFERENCE		4 (bases 1 to 1412)
AUTHORS		Hugenholtz,P.
TITLE		Direct Submission
JOURNAL		Submitted (17-MAY-2000) Microbiology and Parasitology, The University of Queensland, Molecular Biosciences Building 76, Cooper Rd., Brisbane, Queensland 4072, Australia
COMMENT		On Feb 16, 2001 this sequence version replaced gi:871655.
FEATURES		Location/Qualifiers
source		1. .1412
		/organism="uncultured bacterium SBR1108"
		/db_xref="taxon:142194"
		/clone="SBR1108"
		/note="clone obtained from a sequencing batch reactor"
		<l. .>1412
		/product="16S ribosomal RNA"
BASE COUNT	359 a	332 c 467 g 254 t
ORIGIN		
Query Match	70.4%;	Score 17.6; DB 2; Length 1412;











```

CDS
/note="PA5375"
2558..4108
/gene="bet11"
/codon_start=1
/transl_table=11
/product="choline transporter Bet1"
/protein_id="AAG08760.1"
/db_xref="GI:9951698"
/translation="MFFGSTALILVLTAAALILY PDSAGALLCRQEWLSHSEFGWYML
AICGLVFFVAVAFSRFGTLKGEHEKPDYSYGAWCMFLSSGGISLLYFAASEPI
DHLVHPPECPGSPQAQALQOLTFLHGLHCWLYALVGLAVGYFAYRHQPLALRS
ALYPIIGERWKGAGHAGVDCRGIEFTLLGLVTLNGLGALQVSSGLELYTLQMPHSKGT
LLAVILASLVATLAASGVKGRIRLNINILVSSLLLEVLVCGSTLELLNGFVQN
LGDYLDGLVLTDFDIYVTGAGAGKSEWLGWTLFYWAWIWSWAPFVGMFIARISRG
RSRELVCGVLLIPLGFTLAWLSVFGNSALDLMNHGATDLGKALEOPSISYLLLE
HYPLISVIFGLSIFVGLFLTLPADSGSWMLANLSRSGGELDEDAPNWLRLTWSAVVT
LVTIGLFGAGNFAMQTVVVLGAPFSAVLILFLEFGLHKANRDIYEAQVGDRRYQVP
PAPTQPT"
complement(4326..5504)
/gene="PA5376"
CDS
complement(4326..5504)
/gene="PA5376"
/codon_start=1
/transl_table=11
/product="probable ATP-binding component of ABC
transporter."
/protein_id="AAG08761.1"
/db_xref="GI:9951699"
/translation="MTAIFRDHVDVIFGTHTKKALKLIDQGLGRAEILKRTGOVLGVE
DACLDVNRGETCVLMGLSGSGKSSLLRNGINKVSRGKLLIEHGSQVDIANCSPAT
LKAMTRKRIAMVQKFAKMPWLTVAENVGLEMGQRPAAERKLLIDKLELQLSOW
RDKRPDSLGSQMOQVGLARALAMDILLIDPEFSAIDPLIRQQLDELVLVLRQJLH
KTIYVSHDLDEALKIGTRIAIMKDGRIQHGKPEEVLSPADYVVFVAHNTPLNP
LCGOSLMLGLDQCIQNDCEICFDQGRDCWIGLSGSDVLKARGQSNVIDLQWRQGDV
VEKLRRPTLVDSIRMRDALQIRYQTHGKLVLOEQDVRVVGVLGDSLYHALLGNLG
complement(5508..6347)
/gene="PA5377"
CDS
complement(5508..6347)
/gene="PA5377"
/codon_start=1
/transl_table=11
/product="probable permease of ABC transporter"
/protein_id="AAG08762.1"
/db_xref="GI:9951700"
/translation="MFLTDHKLPLGEQIAGFVDWLTLCGASPFDKISDTLEFLIHCVT
NGLLWNPVLIALVGLVHFQIORSGLVFPALASLLILNLGFWQETMETLAQVIFA
TLVCIAGVPLGIVAAKHPWYFALRPLDLMQTVPTFVLYTLFLGLGVVPGLLS
TVIFAVAPRLRLICLGIQDQVPAELLDAGKAFGSRWOLLTRIELPHAMPISAGITQC
IMLSLHMVITAAIVGADGLGKPVYNALNTADISLGFEAGLIAIVLLAILLDRCVKORAV
KGK"
complement(6389..7327)
/gene="PA5378"
CDS
complement(6389..7327)
/gene="PA5378"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG08763.1"
/db_xref="GI:9951701"
/translation="MKGCKSWLLAVSLCAPLAQAAPESCAVRPAQVGVGTDIIVTT
AVTRQVLESIGYTKVKNLLSVPTYSRLNDDVFLGNMPTMNDIKQYAEKGTVE
TLRANLEGAKYTLAVFQYVDGLKRSFADIAKFSKLGKNIYIEPNDGNRAQSMI
DKNAFLGKFLVSEAGMLSQVQRAIRNQWYFLGWEPHMMTRFMQKYLEGGDD
FFPNYGGATITYNVRKGYQAQCANVQGLLKNLSFTELEMKLMDAVLNENKKPEEA
KAWLKDPQDLAWLAGVTTTRDQGPAAAAAKLAFK"
7791..9167
/gene="sdaB"
/note="PA5379"
7791..9167
/gene="sdaB"
/codon_start=1
/transl_table=11

```

```

/product="L-serine dehydratase"
/protein_id="AAG08764.1"
/db_xref="GI:9951702"
/translation="MAISVDFLFIKIGTSSHTVGMRAAALFTALRORQALSARV
RLVFLYSLSATGVGHGTDRVIMGLMGWPDRIPTOIAPRMAALLDGLLQLDGR
VAVPFDVWRDMLDENLPHYPMATLIVAGEOGLHOETIYYSVGGVFDVDAQAASG
RQDQDHVLPYDFSSAEELLCLCKRNLRSVSELMANERLWRSETIREGLRRIWQAM
LDCVQGLRNEGILPGNLVQRRARLRNLQEIIGKPNVIGSTLSAMEWVNFALAVN
EENAGGLMRTAPTNGAAGIIPAVLHYMKFPDQADQDVVDYLLAAAVGILCKKNA
SISGAEEVCGQEGVSACAKAAAGLAEVLGATPAQVENAAEIGLEHNLGLTCDPVGGLV
QVPCIERNAIAAVKAINAAQMALRGDQGHFISLDKVIKTRMDTGADHMDKYEKTSRGG
LAVNAIEC"
9587..10690
/gene="PA5380"
CDS
9587..10690
/gene="PA5380"
/codon_start=1
/transl_table=11
/product="probable transcriptional regulator"
/protein_id="AAG08765.1"
/db_xref="GI:9951703"
/translation="MTTYAPGVPPQNNRPNQSTIGFLLDNFTLISLASAVEPLRMANQL
SGRELYRHTLSLDGQVWASDGLQITPDAGTDNAPVDCVIVCGVGVTQSVTRRHV
TELQAOQRGRRLGAVCTGSMALRAGLIDGYDCSVHWECLAAMQEAEPFVAMSTRLF
SIDNRFTSSGGTAPMDMLHLIGRHGRLSAAISEMFIYERINEODHVRVPLKHM
LGTNQPQLQEIIVALMEANLEEPIDLDLAVVNVSRROLERLFOKYLHCSPSRYLKL
RLIRAROLLKQTSMTIEVASVGVSTPHFSKCYREYFGIPPREROGOPLCQPVL
MPLQDIALMPNSALSALSQAQGSTASVRI"
complement(10737..11114)
/gene="PA5381"
CDS
complement(10737..11114)
/gene="PA5381"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG08766.1"
/db_xref="GI:9951704"
/translation="MAVSLAVELAEIDLPMRSRRDPSLAGSDVGSAAESMAIPIEGAAAM
MHADLIDQDFERLQALGFSPDPSTPEQACEAVRGLSPERAQALRLVDEMLGGH
ATLLPAVREATISROLLPALVPRG"
BASE COUNT 1817 a 3702 c 3673 g 2021 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 1; Length 11213;
Best Local Similarity 83.3%; Pred. No. 7.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cgtcaggtcaggtcagcagcttg 25
Db 943 CGTCCGCTCCCGTGAGCAGCTTG 966

RESULT 43
MTCI376/c LOCUS
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 32/162.
ACCESSION Z95972.1
VERSION Z95972.1
KEYWORDS GI:3261790
SOURCE Mycobacterium tuberculosis.
ORGANISM Mycobacterium tuberculosis.
REFERENCE 1 (bases 1 to 19770)
AUTHORS Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,
Tekalia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,

```

Squares,S., Sqaes,R., Sulston,J.E., Taylor,K., Taylor,K., Whitehead,S. and Barrell,B.G.  
Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence  
NATURE 393 (6685), 537-544 (1998)  
98295987  
Erratum:[published erratum appears in Nature 1998 Nov 12;396(6707):190]]  
12;396(6707):190]]  
2 (bases 1 to 19770)  
Parkhill,J.  
Direct Submission  
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk  
On Jun 27, 1998 this sequence version replaced gi:2143285.

## COMMENT

Notes:  
Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.  
(URL, [http://www.sanger.ac.uk/Projects/M\\_tuberculosis/](http://www.sanger.ac.uk/Projects/M_tuberculosis/)) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.  
Gene prediction was based on a Hidden Markov Model of TB genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.  
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

## FEATURES

Location/Qualifiers  
1..19770  
/organism="Mycobacterium tuberculosis"  
/strain="H37Rv"  
/db\_xref="taxon:1773"  
/clone="I376"  
57..61  
/note="aggaa, possible rbs upstream of Rv0654"  
68..1573  
/gene="Rv0654"  
68..1573  
/gene="Rv0654"  
/note="Rv0654", (MTC1376.22), len: 501, unknown, FASTA score: 053353 LIGNOSTILBENE-ALPHA, BETA-DIOXYGENASE (485 aa) opt:280 z-score: 330.1 E(): 2.3e-11, (28.5% identity in 523 aa overlap). Also similar to M. tuberculosis protein MTCV21C12.07c (29.5% identity in 522 aa overlap)"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein Rv0654"  
/protein\_id="CAB09380.1"  
/db\_xref="GI:2143307"  
/db\_xref="SPTREMBL:O06785"  
/translation="MTAQAEASNPYLEGLAPVSTEVATDLPTVGRIPHELDGRYLRNGPNVAEVDPATYHWFCTGDAMVGVVALRDGARWYRNWRTFAVCAALGEPISARPHPTGIIIEGPNNTVLTAGRTLALVEAGVYNYELTDELTVGCPDFDGTLLHGTYAPQRPDPTGELHVASYSFAGHRVQSVITGDGHARTVDIEVAGSPMHHSFLTDN YVYLDLPVTFDPMQVVPASVPRLVQSVLGRVRIIPDIAALGNRMQGHSDRLPYANPSPYARVGVMPREGGNEVDIEPCYVYHPLNAYSECNGAEVILVDVRYSRMFDRRPGDGRPSLDRWTINLATGATVACRDRAQEFPRINTLTVGGPHR FAYTVIEGGVIEGAGAAALSTPLKQDCVTGSSVTASLDPLDLEGFVFPNPSARAE DDGILMGYGNHGRGDEGQLLLDQALESIAATVHLPPQVPMFGHGNWAPT"  
1585..2664  
/gene="Rv0655"  
1585..2664  
/gene="Rv0655"  
/note="Rv0655", (MTC1376.21), len: 359, abc transporter, FASTA score: YRBE\_ECOLI P45393 hypothetical abc transporter atp-binding (269 aa) opt: 644 z-score: 721.8 E(): 3.4e-33 (38.5% identity in 244 aa overlap); contains

PS00017 ATP/GTP-binding site motif A, PS00211 ABC transporters family signature, highly similar to M. leprae MKL\_MYCLE P30769 possible ribonucleotide transport atp- (347 aa) opt: 2021 z-score:2244.4 E(): 0, (92.2% identity in 335 aa overlap). Also similar to many other M. tuberculosis ABC transporters eg. MTCY253.24 (33.6% identity in 241 aa overlap)"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein Rv0655"  
/protein\_id="CAB09379.1"  
/db\_xref="GI:2143306"  
/db\_xref="SPTREMBL:O06784"  
/translation="MRYSDSVHTTGRWQPRASTGTFPMGVSTEVNGLTKSFGSSRIWE DVTLTIPAGEVSVLLGPGSTGKSVFLKSLIGLRRPGRSIIIDGTDIETCSAKELYEI RTLFGVLQDGAALGSMNLVNDTAPLREHTKKKESEIRDIVMEKLAIVGLGGDEKFF PGEISGMRKRAGLARVALDQPIILCEPDSGLDPVETAYLSQLIMDINAQIDATIL IVTHINARTVPDNNMLFRKHLVMFGRVLLTSDPEVVFQFLNGRRIGPIGMSEE KDEATMAEEQALLDAGHHAGGVEETGVPPOISATPGMPERKAVARQARVREMLHTL PKKAQAAIILDDLEGTHKYAVHEIGQ"  
1762..1785  
/gene="Rv0655"  
/note="PS00017 ATP/GTP-binding site motif A"  
2074..2118  
/gene="Rv0655"  
/note="PS00211 ABC transporters family signature"  
complement(3052..3435)  
/gene="Rv0656c"  
complement(3052..3435)  
/gene="Rv0656c"  
/note="Rv0656c, (MTC1376.20, unknown), len: 127 aa"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein Rv0656c"  
/protein\_id="CAB09378.1"  
/db\_xref="GI:2143305"  
/db\_xref="SPTREMBL:O06783"  
/translation="MAAATTCTGTHRGLELRAAQRAVGCPEPQAEFCRSARNADFDQ MSRMFGDVPDVPKSVWRWIDSQHLRLAAGAVGALSVDLLICDTAAARGLVVLH DDADVELAERHLPDINRVRVVSADD"  
complement(3530..3685)  
/gene="Rv0657c"  
complement(3530..3685)  
/gene="Rv0657c"  
/note="Rv0657c, (MTC1376.19), unknown, len: 51 aa; similar to several other M. tuberculosis hypothetical proteins eg. YW08\_MYCRU Q10848 hypothetical 8.9 kd protein c139.08c (80 aa), fasta scores; opt: 107 z-score: 182.3 E(): 0.0038, 45.8% identity in 48 aa overlap. Also similar to MTCY48.5 andAL020958|SC4H8.7 Streptomyces coelicolor cosmid 4H8 (66 aa), 41.0% identity in 39 aa overlap."  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein Rv0657c"  
/protein\_id="CAB09377.1"  
/db\_xref="GI:2143304"  
/db\_xref="SPTREMBL:O06782"  
/translation="MSVTQIDLDLADVMRIAAVHTTKEAVNLAMRDYVERFRRIE ALARSRE"  
complement(3691..3696)  
/note="possible RBS upstream of Rv0657c"  
complement(3761..4477)  
/gene="Rv0658c"  
complement(3761..4477)  
/note="Rv0658c, (MTC1376.18), len: 238, unknown, problematicmembrane protein, similar to YPRB\_ECOLI P33774 hypothetical124.3 kd protein (urf 1) (217 aa), fasta scores; opt: 174 z-score: 215.8 E(): 5.3e-05, (25.6% identity in 223 aa overlap). Also similar to MTCY359.10 (28.7% identity in 178 aa overlap)"  
/codon\_start=1  
/transl\_table=1



```

/protein_id="CAA70671.1"
/db_xref="GI:2145365"
/trans_id="SPTREMBL:006713"
/translation="MOEITIKLSEAKQOEERKKADLRQKEALQSVNMELETVTDRLT
RGNORTELKQOLQSVTSRDKSCQQAAMALRTPTEEQIKKPKRSEELNVLQ
KNEERNTLVOKTEAENNTIOAYEAVOTVYHLVCTERSLRTMTEARKSOHTLHLQ
REKARVALLITKELAKLITAGKPCPCGSGTDHDPASVHETYEADSHLEDIKRTDVL
TEAALUSQELLSAKMLEQSGARFIEQCPFLQTIQONLEAAASFHQPVYAFETAK
FEWKRIKQILSVKTRMAQMGAYQESLKAQELNKGIFERKEADRIESIISELOSS
MDSLNMFEAFQNVDEAEKQQAIEEKDRAAECEKRIEKSIAFLAEHQAQKEL
RESGHLREKLELHVAAERIKSVIADYEHGLGDYAKGDSIPQLRSVQODLKLKEK
EOSLYELOSAOMKLNQAKSRASASLTIQEAQGRLEKAKAAWLEHTKNTSITRTEEY
EQLSLIPADELEKMKTGIDQPMKLNKMLQT"
4834..5339
/note="putative" - Homology with Sbc from C. perfringens"
/codon_start=1
/transl_table=11
/product="Yisa"
/protein_id="CAA70672.1"
/db_xref="GI:2145366"
/trans_id="SPTREMBL:006714"
/translation="MHIDRLDKLQAVFKNGTFFVEFLAEQLESVARDASARLSMLTRQ
RYAIEVDSEGGFVMDRANGVRRPVSSLSGGTFLTSLSLALSAQQLRGEYPLQ
FFLDREGFGLDQDLDTVYTALEKLSQNLAVGVISHVQELRRLPKKLIVHPAEP
GRGTRYSLELM"
5333..5655
/note="putative"
/codon_start=1
/transl_table=11
/product="Yisb"
/protein_id="CAA70673.1"
/db_xref="GI:2145367"
/trans_id="SPTREMBL:006715"
/translation="MAKQIAGICELGCRSDVQLTEHHLTPKEGGTFLPTAMLCIPCH
KQJHALYTNQELAVRLNGMAELRSDPELAFVWKIRKQPEKLIKTKSNERRKKK"
complement(5692..5919)
/note="putative"
/codon_start=1
/transl_table=11
/product="Yisc"
/protein_id="CAA70674.1"
/db_xref="GI:2145368"
/trans_id="SPTREMBL:006716"
/translation="MSPPAIVGPIAINISGVVNGDSFYLSPKSSKSKALSGSAG
NTGDFLLNNVAVNATNIDPDVNDQDMVGNG"
complement(5942..6343)
/note="putative"
/codon_start=1
/transl_table=11
/product="Yisp"
/protein_id="CAA70675.1"
/db_xref="GI:2145369"
/trans_id="SPTREMBL:006717"
/translation="MLKRIIRLIRKFNLSGIASVFOVQDNTNEIDMSKVYFAVORSLS
TFYHNGSFNKEVOIQOAVKPLPETGVQSAFCHEVPAIVYRSIKIQGVSSAVLH
AGSASLIRGARLAKHQIQSPRSQSPAKNI"
complement(6343..6519)
/note="putative"
/codon_start=1
/transl_table=11
/product="Yisg"
/protein_id="CAA70676.1"
/db_xref="GI:2145370"
/trans_id="SPTREMBL:006718"
/translation="MIFTVINRSLEVDIRMGVSSSVPHIGDTESIYLSIFDTPPP
ESLIIGPPAPLAPE"
complement(6516..7133)
/note="putative"
/codon_start=1
/transl_table=11
/product="Yisf"
/protein_id="CAA70677.1"
/db_xref="GI:2145371"

```

```

/db_xref="SPTREMBL:006719"
/translation="MYDQSVSSYLONLNSFVQQAIIHQOLEROLKEIQTETMNTMKOR
PATTIERVEYRFDQKIERLDGTNLGINLPTDPNSVQNFVSTPQIGMMQDESAQ
LMQOIRQNVMYLTEEIPDILEQLENOYDSRDDTNRHHVIEDIRKQMDSRITYMISH
IKKEENTPPAQAIEHIAEHVKRDIRAVEHFLEHIPSEMGKDEQA"
complement(7156..7389)
/note="putative"
/codon_start=1
/transl_table=11
/product="Yisg"
/protein_id="CAA70678.1"
/db_xref="GI:2145372"
/trans_id="SPTREMBL:006720"
/translation="MNFYINOTIQINYLRLFESINSSILQIGSAGSIKSLNLYNTGS
IYVPAPEVSGSGQPLQLEPDTGSLVLPQPPGR"
complement(7404..7625)
/note="putative"
/codon_start=1
/transl_table=11
/product="Yish"
/protein_id="CAA70679.1"
/db_xref="GI:2145373"
/trans_id="SPTREMBL:006721"
/translation="MPAIVGAFKINAIGTSGVWHIGDCITISPOAQVTRFAGAGSFNT
GDSIKVMYQVATNVYDNDVQPIVANA"
complement(8048..8218)
/note="putative"
/codon_start=1
/transl_table=11

```

Query Match 70.4%; Score 17.6; DB 2; Length 53533;  
 Best Local Similarity 83.3%; Pred. No. 6e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcgtcagggtcaggtgcagcagctt 24  
 || ||||| ||||| ||||| |||||

Db 7542 TCCTCAGGCTCAGGTCAGAACGTT 7519

RESULT 45  
 AC022984/c

LOCUS 72360 bp DNA HTG 13-JUL-2000  
 DEFINITION Homo sapiens chromosome 21 clone RP11-565K18 map 21, LOW-PASS  
 SEQUENCE SAMPLING.

AC022984

AC022984

AC022984.2 GI:9163847

HTG; HTGS\_PHASE0.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 72360)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 21, clone RP11-565K18

Unpublished

REFERENCE

1 (bases 1 to 72360)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,

Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,

Choepe, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P.,

DeArellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,

Ferreira, P., FitzHugh, W., Forrest, C., Gage, B., Galagan, J.,

Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

Landers, T. C., Lechoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K.,

Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,

McPheeters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J.,

Norman, C. H., O'Connor, T., O'Donnell, P., Olivari, T. M., Peterson, K.,

Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,

Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,

Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,

Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,

Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (07-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 13, 2000 this sequence version replaced gi:6921662.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L6502  
Center clone name: 565\_K\_18  
-----

\* NOTE: This record contains 83 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
\* 780 879: gap of 100 bp  
\* 880 1646: contig of 767 bp in length  
\* 1647 1746: gap of 100 bp  
\* 1747 2524: contig of 778 bp in length  
\* 2525 2624: gap of 100 bp  
\* 2625 3385: contig of 761 bp in length  
\* 3386 3485: gap of 100 bp  
\* 3486 4300: contig of 815 bp in length  
\* 4301 4400: gap of 100 bp  
\* 4401 5179: contig of 779 bp in length  
\* 5180 5279: gap of 100 bp  
\* 5280 6039: contig of 760 bp in length  
\* 6040 6139: gap of 100 bp  
\* 6140 6884: contig of 745 bp in length  
\* 6885 6984: gap of 100 bp  
\* 6985 7803: contig of 819 bp in length  
\* 7804 7903: gap of 100 bp  
\* 7904 8696: contig of 793 bp in length  
\* 8697 8796: gap of 100 bp  
\* 8977 9571: contig of 775 bp in length  
\* 9572 9671: gap of 100 bp  
\* 9672 10439: contig of 768 bp in length  
\* 10440 10539: gap of 100 bp  
\* 10540 11318: contig of 779 bp in length  
\* 11319 11418: gap of 100 bp  
\* 11419 12182: contig of 764 bp in length  
\* 12183 12282: gap of 100 bp  
\* 12283 13067: contig of 785 bp in length  
\* 13068 13167: gap of 100 bp  
\* 13168 13946: contig of 773 bp in length  
\* 13941 14040: gap of 100 bp  
\* 14041 14811: contig of 771 bp in length  
\* 14812 14911: gap of 100 bp  
\* 14912 15668: contig of 757 bp in length  
\* 15669 15768: gap of 100 bp  
\* 15769 16516: contig of 748 bp in length  
\* 16517 16616: gap of 100 bp  
\* 16617 17370: contig of 754 bp in length  
\* 17371 17470: gap of 100 bp  
\* 17471 18243: contig of 773 bp in length  
\* 18244 18343: gap of 100 bp  
\* 18344 19109: contig of 766 bp in length  
\* 19110 19209: gap of 100 bp  
\* 19210 19996: contig of 787 bp in length  
\* 19997 20096: gap of 100 bp  
\* 20097 20855: contig of 759 bp in length  
\* 20856 20955: gap of 100 bp  
\* 20956 21780: contig of 825 bp in length  
\* 21781 21880: gap of 100 bp  
\* 21881 22629: contig of 749 bp in length  
\* 22630 22729: gap of 100 bp  
\* 22730 23493: contig of 764 bp in length  
\* 23494 23593: gap of 100 bp  
\* 23594 24375: contig of 782 bp in length  
\* 24376 24475: gap of 100 bp  
\* 24476 25251: contig of 776 bp in length  
\* 25252 25351: gap of 100 bp  
\* 25352 26111: contig of 760 bp in length  
\* 26112 26211: gap of 100 bp  
\* 26212 26953: contig of 742 bp in length  
\* 26954 27053: gap of 100 bp  
\* 27054 27796: contig of 743 bp in length  
\* 27797 27896: gap of 100 bp  
\* 27897 28725: contig of 829 bp in length  
\* 28726 28825: gap of 100 bp  
\* 28826 29604: contig of 779 bp in length  
\* 29605 29704: gap of 100 bp  
\* 29705 30467: contig of 763 bp in length  
\* 30468 30567: gap of 100 bp  
\* 30568 31336: contig of 769 bp in length  
\* 31337 31436: gap of 100 bp  
\* 31437 32208: contig of 772 bp in length  
\* 32209 32308: gap of 100 bp  
\* 32309 33091: contig of 783 bp in length  
\* 33092 33191: gap of 100 bp  
\* 33192 33966: contig of 775 bp in length  
\* 33967 34066: gap of 100 bp  
\* 34067 34837: contig of 771 bp in length  
\* 34838 34937: gap of 100 bp  
\* 34938 35707: contig of 770 bp in length  
\* 35708 35807: gap of 100 bp  
\* 35808 36574: contig of 767 bp in length  
\* 36575 36674: gap of 100 bp  
\* 36675 37440: contig of 766 bp in length  
\* 37441 37540: gap of 100 bp  
\* 37541 38296: contig of 756 bp in length  
\* 38297 38396: gap of 100 bp  
\* 38397 39163: contig of 767 bp in length  
\* 39164 39263: gap of 100 bp  
\* 39264 40040: contig of 777 bp in length  
\* 40041 40140: gap of 100 bp  
\* 40141 40914: contig of 774 bp in length  
\* 40915 41014: gap of 100 bp  
\* 41015 41801: contig of 787 bp in length  
\* 41802 41901: gap of 100 bp  
\* 41902 42675: contig of 774 bp in length  
\* 42676 42775: gap of 100 bp  
\* 42776 43558: contig of 783 bp in length  
\* 43559 43658: gap of 100 bp  
\* 43659 44316: contig of 658 bp in length  
\* 44317 44416: gap of 100 bp  
\* 44417 45180: contig of 764 bp in length  
\* 45181 45280: gap of 100 bp  
\* 45281 46049: contig of 769 bp in length  
\* 46050 46149: gap of 100 bp  
\* 46150 46911: contig of 762 bp in length  
\* 46912 47011: gap of 100 bp  
\* 47012 47839: contig of 828 bp in length  
\* 47840 47939: gap of 100 bp  
\* 47940 48720: contig of 781 bp in length  
\* 48721 48820: gap of 100 bp  
\* 48821 49570: contig of 750 bp in length  
\* 49571 49670: gap of 100 bp  
\* 49671 50454: contig of 784 bp in length  
\* 50455 50554: gap of 100 bp  
\* 50555 51323: contig of 769 bp in length  
\* 51324 51423: gap of 100 bp  
\* 51424 52207: contig of 784 bp in length

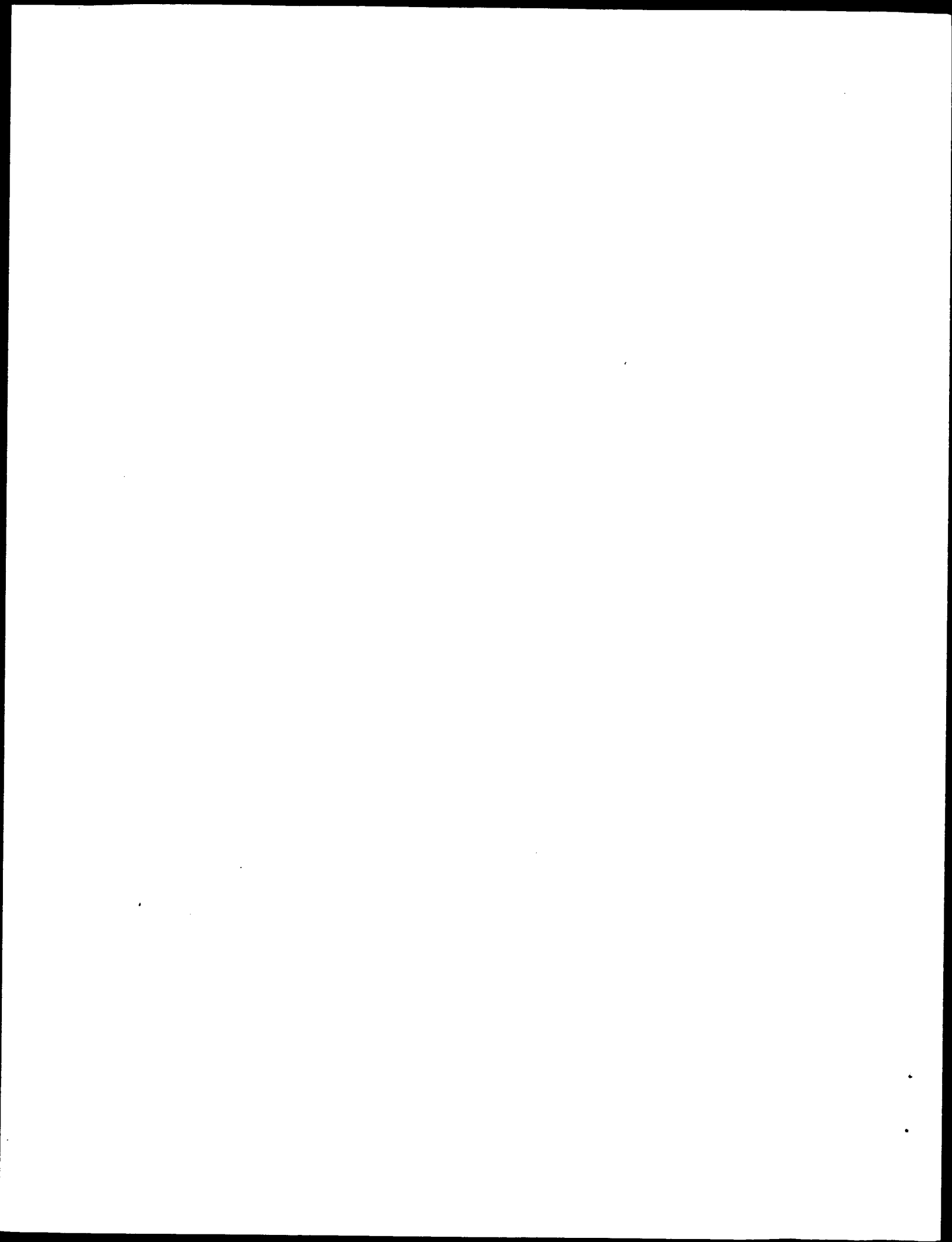


\* 52208 52307: gap of 100 bp  
\* 52308 53069: contig of 762 bp in length  
\* 53070 53169: gap of 100 bp  
\* 53170 53935: contig of 766 bp in length  
\* 53936 54035: gap of 100 bp  
\* 54036 54797: contig of 762 bp in length  
\* 54798 54897: gap of 100 bp  
\* 54898 55813: contig of 916 bp in length  
\* 55814 55913: gap of 100 bp  
\* 55914 56668: contig of 755 bp in length  
\* 56669 56768: gap of 100 bp  
\* 56769 57513: contig of 745 bp in length  
\* 57514 57613: gap of 100 bp  
\* 57614 58389: contig of 776 bp in length  
\* 58390 58489: gap of 100 bp  
\* 58490 59250: contig of 761 bp in length  
\* 59251 59350: gap of 100 bp  
\* 59351 60175: contig of 825 bp in length  
\* 60176 60275: gap of 100 bp  
\* 60276 61057: contig of 782 bp in length  
\* 61058 61157: gap of 100 bp  
\* 61158 61927: contig of 770 bp in length

Query Match 70.4%; Score 17.6; DB 67; Length 72360;  
Best Local Similarity 83.3%; Pred. No. 5.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cgtcaggtgcaggtcacgccttg 25  
||||| ||||| ||||| |||||  
Db 47038 CGTCAGGCGCAGCTCAGATCGTTG 47015

Search completed: October 9, 2001, 12:15:16  
Job time: 4031 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:40:03 ; Search time 470.56 Seconds

(without alignments)  
33.359 Million cell updates/sec

Title: US-09-396-196f-10

Perfect score: 25

Sequence: 1 tctcagggtcaggtcagcagcttg 25

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_0601.\*

1:	/SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT.*
2:	/SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT.*
3:	/SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT.*
4:	/SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT.*
5:	/SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT.*
6:	/SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT.*
7:	/SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT.*
8:	/SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT.*
9:	/SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT.*
10:	/SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.*
11:	/SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT.*
12:	/SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.*
13:	/SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT.*
14:	/SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.*
15:	/SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.*
16:	/SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.*
17:	/SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.*
18:	/SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.*
19:	/SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.*
20:	/SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.*
21:	/SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.*
22:	/SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	20 AAX01303	E. coli biotin syn
2	25	100.0	1084	10 AAN91329	E.coli bio B gene.
3	25	100.0	1121	7 AAN60496	Sequence encoding
4	25	100.0	5872	15 AAO62386	Biotin-biosynthesis
5	17.6	70.4	7286	17 AAT14547	Cytotactin gene.
6	17.6	70.4	7346	20 AAX77924	Human tenascin CDN
7	17.6	70.4	7560	22 AAC83437	Human tenascin-C.
8	17.4	69.6	1526	21 AAZ49706	Mouse growth hormo
9	17.2	68.8	92934	21 AAA81473	N. meningitidis pa
10	17.2	68.8	172325	21 AAF21613	Neisseria meningit
11	17.2	68.8	837096	21 AAA81489	N. meningitidis pa

C 12	17	68.0	1883	14 AAO45959	mRXR-alpha. Mus m
C 13	17	68.0	2095	12 AAO13387	Mouse RXR-alpha co
C 14	17	68.0	3573	17 AAT36507	3.5 kb DNA contg.
C 15	17	68.0	20303	18 AAT71699	Human deoxycytidyl
C 16	17	68.0	26764	18 AAT71696	Human deoxycytidyl
C 17	16.8	67.2	1183	20 AAZ31920	Human helicase, Re
C 18	16.8	67.2	3707	20 AAZ31910	Human helicase, Re
C 19	16.6	66.4	888	22 AAF71212	Corynebacterium gl
C 20	16.6	66.4	888	22 AAF71213	Corynebacterium gl
C 21	16.6	66.4	891	20 AAZ16847	Human gene express
C 22	16.6	66.4	891	20 AAZ16835	Human gene express
C 23	16.6	66.4	1953	21 AAC77412	Human OREF ORF2967
C 24	16.2	64.8	735	19 AAV52499	Streptococcus pneu
C 25	16.2	64.8	1299	18 AAX30776	Streptococcus pneu
C 26	16.2	64.8	1496	8 AAN71388	Encodes subtilisin
C 27	16.2	64.8	2517	20 AAZ28269	Rat neuronal immed
C 28	16.2	64.8	2674	19 AAZ96299	S. pneumoniae derl
C 29	16.2	64.8	2674	19 AAV42984	Streptococcus pneu
C 30	16.2	64.8	10357	19 AAV52324	Streptococcus pneu
C 31	16.2	64.8	49272	19 AAV35000	Mycobacteriophage
C 32	16	64.0	258	21 AAZ46040	partial cDNA encod
C 33	16	64.0	403	21 AAF12855	Aspergillus oryzae
C 34	16	64.0	1188	21 AAC43677	zea mays DNA fragm
C 35	16	64.0	1866	12 AAO13386	Human RXR-alpha co
C 36	16	64.0	2502	18 AAT88991	Streptococcus pneu
C 37	16	64.0	2586	21 AAA79731	Eucalyptus grandis
C 38	16	64.0	4376	21 AAA14818	DNA encoding a Cl
C 39	16	64.0	5837	21 AAA92454	Murine E-selectin
C 40	16	64.0	11770	19 AAV52305	Streptococcus pneu
C 41	16	64.0	50925	21 AAA81487	N. meningitidis pa
C 42	16	64.0	349980	21 AAF21610	Neisseria meningit
C 43	15.8	63.2	244	21 AAC30833	Human secreted pro
C 44	15.8	63.2	549	22 AAF71069	C. glutamicum Str
C 45	15.8	63.2	665	21 AAC43992	Arabidopsis thalia

#### ALIGNMENTS

#### RESULT 1

AAX01303  
ID AAX01303 standard; DNA: 1041 BP.  
XX AC AAX01303;  
XX DT 12-APR-1999 (first entry)  
XX DE E. coli biotin synthetase (BioB) coding sequence.  
XX KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;  
XX KW biotin synthase; biotin production; vitamin H; BioB; ss.  
XX OS Escherichia coli.  
XX PN US5869719-A.  
XX PD 09-FEB-1999.  
XX PF 30-APR-1997; 97US-0846338.  
XX PR 30-APR-1997; 97US-0846338.  
XX PR 08-MAR-1995; 95US-0401068.  
XX (NOVS) NOVARTIS FINANCE CORP.  
XX Patton DA:  
XX WPI; 1999-152902/13.  
XX P-PSDB; AAW73906.

Transgenic plants with high biotin levels - transformed with DNA encoding di:amino-pelargonic acid amino-transferase or biotin synthase

```

XX Example 2; Column 37-40; 34pp; English.
PS
CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC be used in the transgenic plant of the invention. The transgenic plant,
CC plant cell or plant tissue is transformed with a chimeric gene encoding
CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC produces more biotin than a non-transgenic plant, cell or tissue. The
CC plant is used as an improved dietary source of biotin (vitamin H) for
CC humans or animals.
XX
XX Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;
SQ

```

Query Match 100.0%; Score 25; DB 20; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 0.026;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 tcgtcaggtgcaggtcagcagcttg 25
    |||||||
Db 111 tcgtcaggtgcaggtcagcagcttg 135

```

RESULT 2  
ID AAN91329 standard; DNA; 1084 BP.  
XX  
AC AAN91329;  
XX  
DT 15-FEB-1990 (first entry)  
XX  
DE E.coli Bio B gene.  
XX  
KW E.coli; Bio B gene; biotin.  
XX  
OS Escherichia coli.  
XX  
FH Key Location/Qualifiers  
FT CDS 24..1064  
FT /\*tag=a  
XX  
PN GB2216530-A.  
XX  
PD 11-OCT-1989.  
XX  
PF 17-MAR-1989; 89GB-0006210.  
XX  
PR 22-MAR-1988; 88GB-0006804.  
PR 17-MAR-1989; 89GB-0006210.  
XX  
PA (UKAG-) UK MIN. AGRIC. FISH.  
XX  
PI Pearson BM, McKee RA;  
XX  
DR WPI; 1989-295085/41. P-PSDB P91392  
XX  
PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes  
PT - derived from E.coli and capable of replication and expression in other  
PT microorganisms, esp. yeast.  
XX  
PS Table 3; page 33-4; 52pp; English.  
XX  
CC The gene can be used in a plasmid for expression of enzymes of the biotin  
CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae  
CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for  
CC Lactobacillus. Insertion of bio B improves biotin yields in  
CC microorganisms which export biotin, or enables growth in media contg.  
CC little or no biotin of organisms unable to synthesise biotin for their  
CC own use.  
XX  
XX Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;  
SQ

Query Match 100.0%; Score 25; DB 10; Length 1084;  
Best Local Similarity 100.0%; Pred. No. 0.026;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 tcgtcaggtgcaggtcagcagcttg 25
    |||||||
Db 134 tcgtcaggtgcaggtcagcagcttg 158

```

RESULT 3  
ID AAN60496 standard; DNA; 1121 BP.  
XX  
AC AAN60496;  
XX  
DT 17-OCT-1991 (first entry)  
XX  
DE Sequence encoding biotin synthesising enzyme.  
XX  
KW Biotin synthetic enzyme; E.coli; desthiobiotin; ds.  
XX  
FH Key Location/Qualifiers  
FT CDS 42..1082  
FT /\*tag= a  
XX  
PN JP61149091-A.  
XX  
PD 07-JUL-1986.  
XX  
PF 24-DEC-1984; 84JP-0272605.  
XX  
PR 24-DEC-1984; 84JP-0272605.  
XX  
PA (NIPS ) NIPPON SODA KK.  
XX  
DR WPI; 1986-216622/33.  
DR P-PSDB; AAP60536.  
XX  
PT Double stranded DNA encoding biotin synthesising enzyme -  
PT comprises transformed mutant E.coli strain contg. cyclic doubled  
PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.  
XX  
PS Disclosure; Page 534; 23pp; Japanese.  
XX  
CC The sequence may be expressed by a transformed E.coli host, cultured  
CC in a medium containing desthiobiotin.  
XX  
SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;  

Query Match 100.0%; Score 25; DB 7; Length 1121;  
Best Local Similarity 100.0%; Pred. No. 0.027;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 tcgtcaggtgcaggtcagcagcttg 25
    |||||||
Db 152 tcgtcaggtgcaggtcagcagcttg 176

```

RESULT 4  
ID AAQ62386 standard; DNA; 5872 BP.  
XX  
AC AAQ62386;  
XX  
DT 16-NOV-1994 (first entry)  
XX  
DE Biotin-biosynthesis genes contg. plasmid pB030A-15/9.  
XX  
KW Biotin; expression; enterobacteria; vitamin H; synthesis;  
KW plasmid; pB030A-15/9; bioB; bioF; bioC; bioA;  
KW promoter plac; biotin synthase; KAPA synthase;  
KW 8-amino-7-oxononanoate synthase; pimeoyl-CoA; DTB synthase;

KW dethiobiotin synthase; DAPA synthase;  
 KW S-adenosyl-L-methionine; 8-amino-7-oxononanoate aminotransferase;  
 XX Seborrhea; dermatitis; ds.

OS Escherichia coli DSM498.

XX Key Location/Qualifiers  
 FH promoter  
 FT 1..96  
 FT /\*tag= a  
 FT /function= "promoter ptac"  
 FT /evidence= EXPERIMENTAL  
 FT 23..28  
 FT -35\_signal  
 FT /\*tag= b  
 FT /standard\_name= "promoter ptac"  
 FT 45..50  
 FT -10\_signal  
 FT /\*tag= c  
 FT /evidence= EXPERIMENTAL  
 FT /standard\_name= "promoter ptac"  
 FT 105..109  
 FT RBS  
 FT /\*tag= d  
 FT /evidence= EXPERIMENTAL  
 FT /standard\_name= "bioB RBS no. 9"  
 FT 117..1157  
 FT CDS  
 FT /\*tag= e  
 FT /product= "biotin synthase"  
 FT /evidence= EXPERIMENTAL  
 FT /gene= "bioB"  
 FT /number= 1  
 FT 1141..1146  
 FT RBS  
 FT /\*tag= f  
 FT /standard\_name= "bioF RBS"  
 FT 1154..2311  
 FT CDS  
 FT /\*tag= g  
 FT /EC\_number= 2.3.1.47  
 FT /product= "KAPA synthase"  
 FT /evidence= EXPERIMENTAL  
 FT /gene= "bioF"  
 FT /number= 2  
 FT /standard\_name= "8-amino-7-oxononanoate synthase"  
 FT 2284..2288  
 FT RBS  
 FT /\*tag= h  
 FT /standard\_name= "bioC RBS"  
 FT 2295..3050  
 FT CDS  
 FT /\*tag= i  
 FT /function= "involved in pimeloyl-CoA synthesis"  
 FT /product= "protein"  
 FT /gene= "bioC"  
 FT /number= 3  
 FT 3030..3033  
 FT RBS  
 FT /\*tag= j  
 FT /standard\_name= "bioD RBS"  
 FT 3043..3753  
 FT CDS  
 FT /\*tag= k  
 FT /EC\_number= 6.3.3.3  
 FT /product= "DTB synthase"  
 FT /evidence= EXPERIMENTAL  
 FT /gene= "bioD15"  
 FT /number= 4  
 FT 3712..3750  
 FT misc\_RNA  
 FT /standard\_name= "dethiobiotin synthase"  
 FT /\*tag= l  
 FT /note= "bioD15 substitution"  
 FT 3742..3746  
 FT RBS  
 FT /\*tag= m  
 FT /standard\_name= "bioA RBS"  
 FT 3750..5039  
 FT CDS  
 FT /\*tag= n  
 FT /EC\_number= 2.6.1.62  
 FT /product= "DAPA synthase"  
 FT /evidence= EXPERIMENTAL  
 FT /gene= "bioA"  
 FT /number= 5  
 FT /standard\_name= "S-adenosyl-L-methionine; 8-amino-

FT RBS 5088..5093  
 FT /\*tag= o  
 FT /standard\_name= "ORFI RBS"  
 FT 5098..5574  
 FT CDS  
 FT /\*tag= p  
 FT /function= "unknown, involved in biotin synthesis"  
 FT /product= "protein"  
 FT /evidence= EXPERIMENTAL  
 FT /gene= "ORFI"  
 FT /number= 6  
 FT 5583..5644  
 FT terminator  
 FT /\*tag= q  
 FT /standard\_name= "rho-independent transcriptional terminator"  
 FT 5583..5605  
 FT stem\_loop  
 FT /\*tag= r  
 FT 509408023-A.  
 FT XX  
 FT PN 14-APR-1994.  
 FT PD  
 FT XX  
 FT 01-OCT-1993; 93WO-EP02688.  
 FT PF  
 FT XX  
 FT 02-OCT-1992; 92CH-0003124.  
 FT PR  
 FT 15-JUL-1993; 93CH-0002134.  
 FT XX  
 FT (LONZ ) LONZA AG.  
 FT PA  
 FT XX  
 FT Birch O, Brass J, Fuhrmann M, Shaw N;  
 FT PI  
 FT XX  
 FT WPI; 1994-135587/16.  
 FT DR  
 FT P-PSDB; AARS1883, AARS1884, AARS1885, AARS1886, AARS1887, AARS3121.  
 FT DR  
 FT XX  
 FT Biotechnological biotin prodn. using enterobacterial biotin-gene  
 FT - providing vitamin H in high yield  
 FT PT  
 FT XX  
 FT Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.  
 FT PS  
 FT XX  
 FT The sequence is derived from plasmid pB030A-15/9 contg. the  
 FT CC bioB, bioF, bioC, and bioA genes responsible for biosynthesis  
 FT CC of biotin, arranged in a transcription unit. Microorganisms  
 FT CC contg. these DNA fragments or plasmids may be used in the prodn.  
 FT CC of biotin. Biotin (Vitamin H) may prevent seborrhea, dermatitis,  
 FT CC loss of appetite and tiredness.  
 FT XX  
 FT Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;  
 FT SQ  
 FT  
 FT Query Match 100.0%; Score 25; DB 15; Length 5872;  
 FT Best Local Similarity 100.0%; Pred. No. 0.032;  
 FT Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 FT  
 FT QY 1 tcgtcagggtcaggtcagcgttg 25  
 FT Db 227 tcgtcagggtcaggtcagcgttg 251  
 FT ||||||||||||||||||||  
 FT  
 FT RESULT 5  
 FT AAT14547/C  
 FT ID AAT14547 standard; DNA; 7286 BP.  
 FT XX  
 FT AC AAT14547;  
 FT XX  
 FT 21-JUN-1996 (first entry)  
 FT XX  
 FT DT Cytotactin gene.  
 FT DE  
 FT XX  
 FT Cytotactin; neuron; neurite; cell attachment; cell elongation; ss.  
 FT KW  
 FT XX  
 FT Homo sapiens.  
 FT OS  
 FT XX  
 FT Key Location/Qualifiers  
 FT FH

FT CDS 55...6654  
 FT /\*tag= a  
 XX WO9608513-A1.  
 PN  
 XX 21-MAR-1996.  
 PD  
 XX  
 PF 14-SEP-1995; 95WO-US11684.  
 PR 16-SEP-1994; 9AUS-0308359.  
 XX (SCRI ) SCRIPPS RES INST.  
 PA  
 XX Crossin KL, Phillips G, Prieto AL;  
 PI  
 DR WPI; 1996-179904/18.  
 DR P-PSDB; AAR94562.  
 XX  
 PT Cytotactin polypeptide(s), derivs. and antibodies - capable of  
 PT stimulating neuronal cell attachment, neurite out-growth and cell  
 PT elongation  
 XX  
 PS Disclosure: Page 100-110; 159pp; English.  
 XX  
 CC The human cytotactin gene (AAT14547) and chicken cytotactin gene  
 CC (AAT14548) code for multidomain extracellular matrix proteins (AAR94562  
 CC and AAR94563, respectively) capable of stimulating neuronal cell  
 CC attachment, neurite outgrowth and cell elongation. The genes,  
 CC or portions of them, can be inserted into a vector and used for  
 CC prodn. in bacteria of e.g. fusion proteins between GST and CT.  
 XX  
 SQ Sequence 7286 BP; 1830 A; 1926 C; 1999 G; 1531 T; 0 other;

Query Match 70.4%; Score 17.6; DB 17; Length 7286;  
 Best Local Similarity 83.3%; Pred. No. 66;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cgtcaggtgcaggtcagcagcttg 25  
 | ||||| ||||| || |||||  
 Db 2200 CTTTCAGGTGCAGGTAAGTACGTGG 2177

RESULT 6  
 ID AAX77924/c  
 AC AAX77924 standard; cDNA; 7346 BP.  
 XX  
 AC AAX77924;  
 XX  
 DT 16-AUG-1999 (first entry)  
 XX  
 DE Human tenascin cDNA.  
 XX  
 KW Tenascin; antipsoriasis; antiviteligo; anticancer; anti-inflammatory;  
 KW cardiovascular; treatment; disease; depigmentation; albinism; cancer;  
 KW psoriasis; vitiligo; metastasis; melanoma; inflammation; restenosis;  
 KW diagnosis; human; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN DE19750702-A1.  
 XX  
 PD 27-MAY-1999.  
 XX  
 PF 15-NOV-1997; 97DE-1050702.  
 XX  
 PR 15-NOV-1997; 97DE-1050702.  
 XX  
 PA (HMRI ) HOECHST MARION ROUSSEL DEUT GMBH.  
 XX  
 PI Peyman A, Uhlmann E, Weiser C;  
 XX  
 DR WPI; 1999-314075/27.

XX Antisense oligonucleotides that bind to sequences encoding human  
 PT tenascin for treating depigmentation, cancer, inflammation and  
 PT cardiovascular disease  
 XX  
 PS Disclosure: Page 8-14; 18pp; German.  
 XX  
 CC This invention describes novel oligonucleotides with up to 17 optionally  
 CC modified nucleotides (nt), or their salts which are capable of binding  
 CC to a nucleic acid encoding an isoform of human tenascin, or a part of it.  
 CC The oligonucleotides of the invention have antipsoriasis, antiviteligo,  
 CC anticancer, anti-inflammatory and cardiovascular activity. The  
 CC oligonucleotides are used to treat or prevent diseases associated with  
 CC (over)expression of tenascin, particularly depigmentation (albinism,  
 CC psoriasis or vitiligo), cancer or metastases, particularly melanoma,  
 CC inflammation or cardiovascular disease (e.g. restenosis). A preferred  
 CC application is treatment of vitiligo. The oligonucleotides may also be  
 CC used for diagnosis of these diseases.  
 XX  
 SQ Sequence 7346 BP; 1840 A; 1945 C; 2020 G; 1541 T; 0 other;

Query Match 70.4%; Score 17.6; DB 20; Length 7346;  
 Best Local Similarity 83.3%; Pred. No. 66;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cgtcaggtgcaggtcagcagcttg 25  
 | ||||| ||||| || |||||  
 Db 2200 CTTTCAGGTGCAGGTAAGTACGTGG 2177

RESULT 7  
 ID AAC83437/c  
 AC AAC83437 standard; DNA; 7560 BP.  
 XX  
 AC AAC83437;  
 XX  
 DT 27-FEB-2001 (first entry)  
 XX  
 DE Human tenascin-C.  
 XX  
 KW Tenascin-C; axon; dendrite; neurite; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200066628-A1.  
 XX  
 PD 09-NOV-2000.  
 XX  
 PF 28-APR-2000; 2000WO-US11647.  
 XX  
 PR 01-MAY-1999; 99US-0132137.  
 XX  
 PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.  
 XX  
 P1 Geller HM, Meiners S;  
 XX  
 DR WPI; 2001-015974/02.  
 DR P-PSDB; AAB36935.  
 XX  
 PT New peptides and methods of stimulating axonal and/or dendritic growth  
 PT and/or guidance by administering a peptide, particularly comprising a  
 PT tenascin-C region, to a neuron  
 XX  
 PS Disclosure: Page 11-16; 73pp; English.  
 XX  
 CC The present invention relates to a tenascin-C peptide. The peptide,  
 CC particularly from a tenascin-C region, is used for stimulating axonal  
 CC and/or dendritic growth, independent of neurite guidance, and  
 CC stimulating axonal and/or dendritic guidance independent of axonal  
 CC and/or dendritic growth.  
 XX  
 SQ Sequence 7560 BP; 1894 A; 2019 C; 2062 G; 1585 T; 0 other;



Query Match 68.8%; Score 17.2; DB 21; Length 92934;  
 Best Local Similarity 86.4%; Pred. No. 1.3e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 tcagggtcagggtcagcagcttg 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 60036 TCAGGTGCAATTCACACGTTG 60015

RESULT 10  
 ID AAF21613 standard; DNA; 172325 BP.  
 AC AAF21613;  
 XX  
 DT 13-MAR-2001 (first entry)  
 XX  
 DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:114.  
 XX  
 DE Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;  
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
 KW ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO200066791-A1.  
 PD 09-NOV-2000.  
 XX  
 PF 08-MAR-2000; 2000WO-US05928.  
 XX  
 PR 30-APR-1999; 99US-0132068.  
 PR 08-OCT-1999; 99WO-US23573.  
 PR 28-FEB-2000; 2000GB-0004695.  
 XX  
 XX (CHIR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;  
 PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;  
 PI Frazer CM, Grandi G;  
 XX  
 DR WPI; 2000-647603/62.

Neisseria meningitidis B full length genome sequence and open reading  
 frames are used to detect, treat and prevent Neisserial infections -  
 Claim 7; Appendix A; 692pp; English.

The present invention describes the full length genome of  
 Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607  
 to AAF21613 represent fragments of the NMB genomic sequence, as the  
 sequence was too long to go in a record on its own it was split into 8  
 sequences which overlap each other at the beginning and end of each  
 sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at  
 the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at  
 the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the  
 Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to  
 AAF21606 represent PCR primers which are used in the exemplification of  
 the present invention. The NMB genome and fragments from it have  
 antibacterial activity, and can be used in vaccines and gene therapy.  
 Neisseria nucleic acids, proteins and/or antibodies which binds to the  
 proteins can be used in compositions for treating or preventing infection  
 due to Neisserial bacteria or as a diagnostic reagent for detecting the  
 presence of Neisserial bacteria or of antibodies raised to Neisserial  
 bacteria. Computers, computer memory, computer storage medium or computer  
 databases can be used in a search to identify open reading frames (ORFs)  
 or coding sequences within the NMB genome. The DNA sequences provide  
 further opportunities to find antigenic or immunogenic proteins which are  
 more effective in vaccines than the outer membrane proteins currently  
 used.

SQ Sequence 172325 BP; 43072 A; 47583 C; 41465 G; 40205 T; 0 other;

Query Match 68.8%; Score 17.2; DB 21; Length 172325;  
 Best Local Similarity 86.4%; Pred. No. 1.4e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 tcagggtcagggtcagcagcttg 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 163703 TCAGGTGCAATTCACACGTTG 163682

RESULT 11  
 ID AAA81489/C  
 XX  
 AC AAA81489 standard; DNA; 837096 BP.  
 XX  
 DT 04-DEC-2000 (first entry)  
 XX  
 DE N. meningitidis partial DNA sequence gnm\_37 SEQ ID NO:37.  
 XX  
 DE Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; MenB; ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO200022430-A2.  
 PD 20-APR-2000.  
 XX  
 PF 08-OCT-1999; 99WO-US23573.  
 XX  
 PR 09-OCT-1998; 98US-0103794.  
 PR 30-APR-1999; 99US-0132068.  
 XX  
 XX (CHIR ) CHIRON CORP.  
 PA  
 XX  
 PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;  
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
 PI Rappuoli R, Pizza M;  
 XX  
 DR WPI; 2000-318079/27.

Isolated nucleotide sequences of Neisseria meningitidis which can be  
 used in the diagnosis and treatment of N. meningitidis infection and  
 other Neisserial infections, for example, N.gonorrhoea -  
 Claim 7; Page 629-865; 1760pp; English.

The present invention describes methods of obtaining immunogenic  
 proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
 represent specifically claimed Neisseria meningitidis genomic DNA  
 sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
 Neisseria DNA sequences and their corresponding proteins; AAA81254 to  
 AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 isolation of Neisseria meningitidis DNA sequences; and AAA81322 to  
 AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
 sequences, which are all used in the exemplification of the present  
 invention. The nucleic acid sequences, protein sequences, and antibodies  
 against them, can be used in the manufacture of a composition. The  
 composition can be used as a medicament (or in the manufacture of a  
 medicament) for treating, preventing or diagnosing infection due to  
 Neisserial bacteria. For example, some of the identified proteins could  
 be components of vaccines against Meningococcus B; against all serotypes;  
 and/or against all pathogenic Neisseriae. Identification of sequences  
 from the bacterium will also facilitate production of biological probes,  
 particularly organism-specific probes. Attempts to make efficacious  
 Meningococcus B vaccines have failed mainly due to antigen tolerance.  
 Multivalent vaccines have also been tried but none have successfully  
 overcome antigenic variability. The provision of further, complete  
 sequences may provide an opportunity to identify secreted or surface



CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.

SQ Sequence 837096 BP; 207534 A; 227065 C; 205215 G; 197280 T; 2 other;

Query Match 68.8%; Score 17.2; DB 21; Length 837096;  
 Best Local Similarity 86.4%; Pred. No. 1.5e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 tcaggtgcaggtcagcagcttg 25  
 ||||| ||||| ||||| |||||  
 DB 828474 TCAGGTGCATTCAACACGTTG 828453

RESULT 12  
 AAQ45959/C  
 ID AAQ45959 standard; DNA; 1883 BP.

XX AC AAQ45959;

XX DT 01-FEB-1994 (first entry)

XX DE mRXR-alpha.

XX KW Retinoid X receptor; RXR; retinoic acid receptor; RAR;  
 KW thyroid receptor; TR; retinoic acid response elements; RARE;  
 KW thyroid receptor response element; TRE;  
 KW retinoid X receptor response element; RXRE;  
 KW subunit; heterodimer; ss.

XX OS Mus musculus.

XX FH Key Location/Qualifiers  
 XX CDS 96..1499  
 XX FT /\*tag= a  
 XX FT /label= RXR-beta

XX PN WO9315216-A.

XX PD 05-AUG-1993.

XX PF 25-JAN-1993; 93WO-US00639.

XX PR 24-JAN-1992; 92US-0825667.

XX (CNRS ) CENT NAT RECH SCI.  
 PA (INRM ) INST NAT SANTE & RECH MEDICALE.  
 PA (SQUIL ) SQUILBB & SONS INC E. R.  
 PA (UYST-) UNIV PASTEUR STRASBOURG LOUIS.

XX PI Chabon P, Kastner P, Leid M;

XX WPI; 1993-258691/32.

XX P-PSDB; AAR39470.

XX Heterodimer comprising retinoic acid or thyroid receptor and  
 PT retinoid X receptor - useful in diagnosis and for generation of  
 PT antibodies

XX Claim 25; Page 97-99; 128pp; English.

XX CC Two different types of nuclear receptors, RAR and TR dimerise with  
 CC RXR to form a heterodimer which is capable of binding to RARE, TRE  
 CC or RXRE at physiological conditions. Sequences encoding 1 isoform  
 CC of mouse RXR-beta, 3 isoforms of human RXR-beta, 1 isoform of mouse  
 CC RXR-alpha and 1 isoform of mouse RXR-gamma are provided (AAQ45955-60).

XX SQ Sequence 1883 BP; 380 A; 595 C; 518 G; 390 T; 0 other;

Query Match 68.0%; Score 17; DB 14; Length 1883;

Best Local Similarity 80.0%; Pred. No. 1.1e+02;  
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 tcgtcaggtcaggtcagcagcttg 25  
 ||| ||||| ||||| |||||  
 DB 626 TCGGCAGGTGTAGGTCAGGCTCTTG 602

RESULT 13  
 AAQ13387/C  
 ID AAQ13387 standard; cDNA; 2095 BP.

XX AC AAQ13387;

XX DT 07-NOV-1991 (first entry)

XX DE Mouse RXR-alpha coding sequence.

XX KW ligand-binding domain receptor-alpha; retinoid receptor;  
 KW vitamin A metabolite; hRAR-alpha; ss.

XX OS Mus musculus.

XX FH Key Location/Qualifiers  
 XX CDS 177..1581  
 XX FT /\*tag= a  
 XX FT /product= mRXR-alpha

XX PN WO9112258-A.

XX PD 22-AUG-1991.

XX PF 22-JAN-1991; 91WO-US00399.

XX PR 09-FEB-1990; 90US-0478071.

XX (SALK ) SALK INST FOR BIOL STUD.

XX PI Mangelsdorf DJ, Evans RM;

XX WPI; 1991-267086/36.

XX P-PSDB; AAR13799.

XX Novel DNA sequence encoding a receptor polypeptide - which is  
 PT useful for testing the ability of cpds. to regulate their  
 PT transcription - activation effects

XX Claim 4; Page 29-30; 56pp; English.

XX CC mRXR-alpha is a novel retinoid receptor. Clones were isolated from a  
 CC lambda gt10 human liver cDNA library by screening with the known  
 CC hRAR-alpha DNA binding domain. These positive clones were subcloned  
 CC and a unique receptor-like sequence was identified and used to  
 CC screen a lambda gt10 human kidney cDNA library. A human RXR-alpha  
 CC coding sequence was identified and this was used to identify  
 CC mRXR-alpha from a mouse embryo library. Labelled single-stranded  
 CC nucleic acid sequences are also claimed comprising 20 or more  
 CC contiguous bases from within bases 20-2095 of the mRXR-alpha  
 CC sequence. These can be used to identify other DNAs encoding  
 CC retinoid receptors. See also AAQ13386 and AAQ13388.

XX SQ Sequence 2095 BP; 404 A; 680 C; 598 G; 413 T; 0 other;

Query Match 68.0%; Score 17; DB 12; Length 2095;  
 Best Local Similarity 80.0%; Pred. No. 1.1e+02;  
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 tcgtcaggtcaggtcagcagcttg 25  
 ||| ||||| ||||| |||||  
 DB 708 TCGGCAGGTGTAGGTCAGGCTCTTG 684

RESULT 14  
AAT36507/c  
ID AAT36507 standard; DNA; 3573 BP.

XX  
XX  
AC AAT36507;

XX  
DT 04-OCT-1996 (first entry)

XX  
DE 3.5 kb DNA contg. ars region, rpmH, dnaA and dnaN genes.

XX  
KW ars; dnaA; dnaN; rpmH; Brevibacterium flavum; low expression;  
XX autonomous replication sequence; plasmid copy number; ds.

XX  
OS Brevibacterium flavum.

XX  
FH Location/Qualifiers  
FT complement (61..84)

FT /\*tag= a

FT /note= "start of rpmH gene"

FT 932..2506

FT /\*tag= b

FT /product= dnaA gene product

FT 3030

FT /\*tag= c

FT /note= "partial dnaN gene product"

FT

XX JP08089259-A.

XX  
PN  
XX  
PD 09-APR-1996.

XX  
PF 05-DEC-1994; 94JP-0301140.

XX  
XX 26-JUL-1994; 94JP-0174316.

XX  
PR (MITU ) MITSUBISHI CHEM CORP.

XX  
XX WPI: 1996-233351/24.

XX  
DR P-FSDB; AAR98522-24.

XX  
XX DNA fragment contg. an autonomous replication sequence - useful for  
XX controlling the copy number of vectors used in cloning

XX  
PS Example 1; Page 11-14; 17pp; Japanese.

XX  
CC The present sequence is a 3.5 kb DNA fragment contg. an autonomous  
XX replicating sequence (ars) isolated from the coryneform bacterium,  
XX Brevibacterium flavum. The 3.5 kb BamHI-SalI fragment contains a partial  
XX rpmH gene, the dnaA gene and partial dnaN gene. The ars region was shown  
XX to be present between the dnaA and rpmH genes and between the dnaA and  
XX dnaN genes. Vectors contg. the ars region have a low copy number which  
XX is useful for cloning of genes which adversely affect cells with high  
XX expression.

XX  
SQ Sequence 3573 BP; 932 A; 941 C; 872 G; 828 T; 0 other;

Query Match 68.0%; Score 17; DB 17; Length 3573;

Best Local Similarity 80.0%; Pred. No. 1.1e+02;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 tcgtcaggtcaggtcagcagcttg 25

Db 1064 TCGTCAGGTCAAAATATGCAGCTTG 1040

RESULT 15

AAT71699/c

ID AAT71699 standard; DNA; 20303 BP.

XX  
XX  
AC AAT71699;

XX  
DT 20-AUG-1997 (first entry)

XX

DE Human deoxycytidylate deaminase intron 2 encoding DNA.

XX Recombinant deaminase; dCMP; ds.

XX  
XX Homo sapiens.

XX  
XX US5622851-A.

XX  
PD 22-APR-1997.

XX  
PF 10-JAN-1995; 95US-0370975.

XX  
PR 10-JAN-1995; 95US-0370975.

XX  
XX (HEAL-) HEALTH RES INC.

XX  
XX Maley F, Maley GR, Weiner KXB;

XX  
XX WPI: 1997-244391/22.

XX  
XX DNA encoding human deoxycytidylate deaminase - for production of  
XX recombinant deaminase

XX  
PS Claim 2; Column 83-100; 58pp; English.

XX  
CC The present sequence encodes the human deoxycytidylate (dCMP)

CC deaminase intron 2, which comprises 20303 base pairs from nucleotides

CC 1964-22266 of the dCMP deaminase sense strand. The dCMP deaminase gene

CC contains a 5' untranslated region (including the promoter), 5 exons,

CC 4 introns and a 3' untranslated region (including the stop signals).

CC The gene can be used to produce recombinant dCMP deaminase, which can

CC be used to convert dCMP to dUMP. Also, the dCMP gene can be altered

CC (removed or mutated) to alter DNA replication in cells, which may lead

CC to mutagenesis.

XX  
SQ Sequence 20303 BP; 5454 A; 4115 C; 5052 G; 5682 T; 0 other;

Query Match 68.0%; Score 17; DB 18; Length 20303;

Best Local Similarity 80.0%; Pred. No. 1.4e+02;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 tcgtcaggtcaggtcagcagcttg 25

Db 285 TGGTGTGGTGCAGGCCAGCAGGTG 261

RESULT 16

AAT71696/c

ID AAT71696 standard; DNA; 26764 BP.

XX  
AC AAT71696;

XX  
DT 20-AUG-1997 (first entry)

XX  
DE Human deoxycytidylate deaminase gene.

XX  
KW Recombinant deaminase; dCMP; ss.

XX  
OS Homo sapiens.

XX  
FH Key Location/Qualifiers

FT misc\_feature 1..1317

FT /\*tag= a

FT /note= "5' untranslated region, including promotor"

FT 1318..1425

FT /\*tag= b

FT /number= 1

FT 1426..1827

FT /\*tag= c

FT /number= 1

FT 1828..1963

FT /\*tag= d

FT  
FT exon

```

FT      intron      /number= 2
FT      1964..22266
FT      /*tag= e
FT      /number= 2
FT      22267..22383
FT      /*tag= f
FT      /number= 3
FT      22384..23740
FT      /*tag= g
FT      /number= 3
FT      23741..23837
FT      /*tag= h
FT      /number= 4
FT      23838..25391
FT      /*tag= i
FT      /number= 4
FT      25392..25467
FT      /*tag= j
FT      /number= 5
FT      25468..26764
FT      /*tag= k
FT      /note= "3' untranslated region"
FT      XX
FT      XX
PN      US5622851-A.
XX
XX      22-APR-1997.
XX
XX      10-JAN-1995; 95US-0370975.
XX
XX      10-JAN-1995; 95US-0370975.
XX
XX      (HEAL-) HEALTH RES INC.
XX
XX      Maley F, Maley GR, Weiner KXB;
XX
XX      WPI: 1997-244391/22.
XX      P-PSDB; AAW18205.
XX
XX      DNA encoding human deoxycytidylate deaminase - for production of
XX      recombinant deaminase
XX
XX      Claim 3; Column 55-78; 58pp; English.
XX
XX      The present sequence encodes the human deoxycytidylate (dCMP)
XX      deaminase gene, which contains a 5' untranslated region (including
XX      the promoter), 5 exons, 4 introns and a 3' untranslated region
XX      (including the stop signals). The gene can be used to produce
XX      recombinant dCMP deaminase, which can be used to convert dCMP to dUMP.
XX      Also, the dCMP gene can be altered (removed or mutated) to alter DNA
XX      replication in cells, which may lead to mutagenesis.
XX
XX      Sequence 26764 BP; 7079 A; 5521 C; 6539 G; 7625 T; 0 other;

Query Match      68.0%; Score 17; DB 18; Length 26764;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 tcgtcaggtgcaggtcagcagttg 25
        ||| ||||| ||||| |||||
DB      2248 TGGTGTGGTGCAGGCAGCAGGTTG 2224

RESULT 17
AAZ31920/C
ID      AAZ31920 standard; cDNA to mRNA; 1183 BP.
XX
XX      AAZ31920;
XX
XX      25-JAN-2000 (first entry)
XX
XX      Human helicase, RecQ5, coding sequence fragment.
XX

KW      Helicase; RecQ5; human; homeostasis; cell aging; diagnosis; ss.
XX
XX      Homo sapiens.
XX
XX      JP11276173-A.
XX
XX      12-OCT-1999.
XX
XX      27-MAR-1998; 98JP-0081492.
XX
XX      27-MAR-1998; 98JP-0081492.
XX
XX      (EIJ-) EIJIN KENKYUSHO KK.
XX
XX      WPI: 1999-626937/54.
XX
XX      A gene coding helicase, RecQ5 - useful in research for elucidating the
XX      relationship to the maintenance of human homeostasis and cell aging
XX
XX      Example 1; Page 20; 25pp; Japanese.
XX
XX      This sequence is a fragment of the gene encoding the human helicase of
XX      the invention, designated RecQ5. The helicase RecQ5 gene can be used in a
XX      method for the diagnosis of diseases caused by an abnormality in the
XX      helicase gene. The gene can be used in researching the relationship
XX      between the maintenance of human homeostasis and cell aging.
XX
XX      Sequence 1183 BP; 258 A; 313 C; 324 G; 288 T; 0 other;

Query Match      67.2%; Score 16.8; DB 20; Length 1183;
Best Local Similarity 90.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 tcgtcaggtgcaggtcagca 20
        ||| ||||| ||||| |||||
DB      713 TCTTCAGGTGCAGGCAGCA 694

RESULT 18
AAZ31910/C
ID      AAZ31910 standard; cDNA; 3707 BP.
XX
XX      AAZ31910;
XX
XX      25-JAN-2000 (first entry)
XX
XX      Human helicase, RecQ5, coding sequence.
XX
XX      Helicase; RecQ5; human; homeostasis; cell aging; diagnosis; ss.
XX
XX      Homo sapiens.
XX
XX      JP11276173-A.
XX
XX      12-OCT-1999.
XX
XX      27-MAR-1998; 98JP-0081492.
XX
XX      27-MAR-1998; 98JP-0081492.
XX
XX      (EIJ-) EIJIN KENKYUSHO KK.
XX
XX      WPI: 1999-626937/54.
XX      P-PSDB; AAY43327.
XX
XX      A gene coding helicase, RecQ5 - useful in research for elucidating the
XX      relationship to the maintenance of human homeostasis and cell aging
XX
XX      Claim 2; Page 13-15; 25pp; Japanese.
XX
XX      This sequence encodes the human helicase of the invention, designated
XX      RecQ5. The helicase RecQ5 gene can be used in a method for the diagnosis

```

CC of diseases caused by an abnormality in the helicase gene. The gene can  
 CC be used in researching the relationship between the maintenance of human  
 CC homeostasis and cell aging.

SQ Sequence 3707 BP; 801 A; 915 C; 1021 G; 970 T; 0 other;

Query Match 67.2%; Score 16.8; DB 20; Length 3707;  
 Best Local Similarity 90.0%; Pred. No. 1.4e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagca 20  
 || |||||  
 Db 791 TCCTCAGGTGCAGGCAGCA 772

## RESULT 19

AAF71212

ID AAF71212 standard; DNA; 888 BP.

XX AAF71212;

AC AAF71212;

DT 30-APR-2001 (first entry)

XX Corynebacterium glutamicum HA protein nucleotide sequence SEQ ID NO:149.

KW Corynebacterium glutamicum; homeostasis; adaptation; HA protein;  
 KW fine chemical production; organic acid; proteinogenic amino acid;  
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;  
 KW diagnosis; Corynebacterium diphtheriae; genetic engineering;  
 KW Brevibacterium; environmental condition; ds.

XX Corynebacterium glutamicum.

OS WO200100842-A2.

PN 04-JAN-2001.

XX 23-JUN-2000; 2000WO-IB00911.

XX 25-JUN-1999; 99US-0141031.

PR 08-JUL-1999; 99DE-1031636.

PR 09-JUL-1999; 99DE-1032125.

PR 09-JUL-1999; 99DE-1032126.

PR 09-JUL-1999; 99DE-1032127.

PR 09-JUL-1999; 99DE-1032128.

PR 09-JUL-1999; 99DE-1032129.

PR 09-JUL-1999; 99DE-1032226.

PR 14-JUL-1999; 99DE-1032920.

PR 14-JUL-1999; 99DE-1032922.

PR 14-JUL-1999; 99DE-1032924.

PR 14-JUL-1999; 99DE-1032928.

PR 14-JUL-1999; 99DE-1032930.

PR 14-JUL-1999; 99DE-1032933.

PR 14-JUL-1999; 99DE-1032935.

PR 14-JUL-1999; 99DE-1032973.

PR 14-JUL-1999; 99DE-1033002.

PR 14-JUL-1999; 99DE-1033003.

PR 14-JUL-1999; 99DE-1033005.

PR 31-AUG-1999; 99DE-1033006.

PR 31-AUG-1999; 99DE-1041378.

PR 31-AUG-1999; 99DE-1041379.

PR 31-AUG-1999; 99DE-1041390.

PR 31-AUG-1999; 99DE-1041391.

PR 03-SEP-1999; 99DE-1042088.

XX (BADI ) BASF AG.

XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

XX WPI; 2001-061974/07.

DR

DR P-PSDB; AAB79097.

XX New isolated Corynebacterium glutamicum nucleic acid for production or  
 PT modulation of production of fine chemicals such as amino acids,  
 PT nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins  
 PT or enzymes -

PS Claim 3; Page 353-354; 712pp; English.

XX AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis  
 CC and adaptation (HA) proteins given in AAB79023 to AAB79242. The  
 CC C. glutamicum HA genes (I) can be used in vectors for expression in host  
 CC cells and production of fine chemicals, such as, an organic acid,  
 CC proteinogenic or nonproteinogenic amino acid (preferred), purine or  
 CC pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated  
 CC fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,  
 CC polyketide or enzyme. The amino acids produced can be lysine, glutamine,  
 CC glutamate, alanine, aspartate, glycine, serine, threonine, methionine,  
 CC cysteine, valine, leucine, isoleucine, arginine, proline, histidine,  
 CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can  
 CC be modulated. The presence of (I) or HA proteins encoded by them are  
 CC used for diagnosing the presence or activity of Corynebacterium  
 CC diphtheriae. (I) can be used to map the C. glutamicum genome or can be  
 CC used as markers for genetically engineered Corynebacterium or  
 CC Brevibacterium. The HA proteins encoded by the (I) are used to maintain  
 CC homeostasis in C. glutamicum or help the microorganism to adapt to  
 CC different environmental conditions.

XX Sequence 888 BP; 229 A; 217 C; 244 G; 198 T; 0 other;

Query Match 66.4%; Score 16.6; DB 22; Length 888;

Best Local Similarity 82.6%; Pred. No. 1.5e+02;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 gtcaggtgcaggtcagcagttg 25

| ||||| || || |||||

Db 652 gccagtgaggctctgcagttg 674

## RESULT 20

AAF71213

ID AAF71213 standard; DNA; 888 BP.

XX AAF71213;

XX AAF71213;

DT 30-APR-2001 (first entry)

XX Corynebacterium glutamicum HA protein nucleotide sequence SEQ ID NO:151.

KW Corynebacterium glutamicum; homeostasis; adaptation; HA protein;  
 KW fine chemical production; organic acid; proteinogenic amino acid;  
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;  
 KW diagnosis; Corynebacterium diphtheriae; genetic engineering;  
 KW Brevibacterium; environmental condition; ds.

OS Corynebacterium glutamicum.

XX WO200100842-A2.

PN 04-JAN-2001.

XX 23-JUN-2000; 2000WO-IB00911.

XX 25-JUN-1999; 99US-0141031.

PR 08-JUL-1999; 99DE-1031636.

PR 09-JUL-1999; 99DE-1032125.

PR 09-JUL-1999; 99DE-1032126.

PR 09-JUL-1999; 99DE-1032127.

PR 09-JUL-1999; 99DE-1032128.

PR 09-JUL-1999; 99DE-1032129.

PR 09-JUL-1999; 99DE-1032226.  
 PR 14-JUL-1999; 99DE-1032920.  
 PR 14-JUL-1999; 99DE-1032922.  
 PR 14-JUL-1999; 99DE-1032924.  
 PR 14-JUL-1999; 99DE-1032928.  
 PR 14-JUL-1999; 99DE-1032930.  
 PR 14-JUL-1999; 99DE-1032933.  
 PR 14-JUL-1999; 99DE-1032935.  
 PR 14-JUL-1999; 99DE-1033002.  
 PR 14-JUL-1999; 99DE-1033003.  
 PR 14-JUL-1999; 99DE-1033005.  
 PR 14-JUL-1999; 99DE-1033006.  
 PR 31-AUG-1999; 99DE-1041378.  
 PR 31-AUG-1999; 99DE-1041379.  
 PR 31-AUG-1999; 99DE-1041390.  
 PR 31-AUG-1999; 99DE-1041391.  
 PR 03-SEP-1999; 99DE-1042088.  
 XX (BADI ) BASF AG.  
 PA  
 XX  
 XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
 PI P-PSDB; AAB79098.  
 DR WPI; 2001-061974/07.  
 DR P-PSDB; AAB79098.  
 XX  
 PT New isolated Corynebacterium glutamicum nucleic acid for production or  
 PT modulation of production of fine chemicals such as amino acids,  
 PT nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins  
 PT or enzymes -  
 XX  
 XX Claim 3; Page 355-356; 712pp; English.  
 PS  
 XX AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis  
 CC and adaptation (HA) proteins given in AAB79023 to AAB79242. The  
 CC C. glutamicum HA genes (I) can be used in vectors for expression in host  
 CC cells and production of fine chemicals, such as, an organic acid,  
 CC proteinogenic or nonproteinogenic amino acid (preferred), purine or  
 CC pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated  
 CC fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,  
 CC polyketide or enzyme. The amino acids produced can be lysine, glutamine,  
 CC glutamate, alanine, aspartate, glycine, serine, threonine, methionine,  
 CC cysteine, valine, leucine, isoleucine, arginine, proline, histidine,  
 CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can  
 CC be modulated. The presence of (I) or HA proteins encoded by then are  
 CC used for diagnosing the presence or activity of Corynebacterium  
 CC diphtheriae. (I) can be used to map the C. glutamicum genome or can be  
 CC used as markers for genetically engineered Corynebacterium or  
 CC Brevibacterium. The HA proteins encoded by the (I) are used to maintain  
 CC homeostasis in C. glutamicum or help the microorganism to adapt to  
 CC different environmental conditions.  
 XX  
 SQ Sequence 888 BP; 229 A; 217 C; 244 G; 198 T; 0 other;  
 Query Match 66.4%; Score 16.6; DB 22; Length 888;  
 Best Local Similarity 82.6%; Pred. No. 1.5e+02;  
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 3 gtcagggtcagggtcagcgttg 25  
 | ||||| || || ||||| |||||  
 Db 652 gccagggtgagttctgcagcgttg 674  
 RESULT 21  
 AAZ16847/C  
 ID AAZ16847 standard; cDNA; 891 BP.  
 XX  
 AC AAZ16847;  
 XX  
 XX 12-OCT-1999 (first entry)  
 DT  
 XX Human gene expression product cDNA sequence SEQ ID NO:4317.  
 DE

XX Human; gene; gene expression product; diagnosis; therapy; probe;  
 KW detection; mapping; tissue typing; profiling; forensic; cancer;  
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9938972-A2.  
 XX  
 PD 05-AUG-1999.  
 XX  
 XX 28-JAN-1999; 99WO-US01619.  
 PF  
 XX 03-APR-1998; 98US-0080666.  
 PR 28-JAN-1998; 98US-0072910.  
 PR 24-FEB-1998; 98US-0075954.  
 PR 31-MAR-1998; 98US-0080114.  
 PR 03-APR-1998; 98US-0080515.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;  
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;  
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;  
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;  
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;  
 XX  
 XX WPI; 1999-494092/41.  
 DR  
 XX Novel human genes and their expression products which are  
 PT differentially expressed in different cell types  
 PT  
 XX Claim 1; Page 2044; 2479pp; English.  
 PS  
 XX The present invention describes a library of human polynucleotides  
 CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is  
 CC a method of detecting differentially expressed genes correlated with the  
 CC cancerous state of a mammalian cell, comprising detecting at least one  
 CC differentially expressed gene product in a test sample from a cell  
 CC suspected of being cancerous, where the gene product is encoded by one  
 CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The  
 CC polynucleotides can be used as a source of primers and probes, which can  
 CC be used for a variety of purpose, e.g. detection of expression levels,  
 CC mapping, tissue typing or profiling, forensics, genetic analysis and  
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
 CC can be used for raising antibodies for experimental, diagnostic and  
 CC therapeutic purposes. The polynucleotides may also be used to construct  
 CC arrays for diagnostics (which may be used to determine function of an  
 CC encoded protein); and to detect differences in expression levels between  
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
 CC identify a genetic predisposition or susceptibility to a disease such as  
 CC cancer). The polynucleotides of the invention are especially used in the  
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
 CC and lung cancer. The polynucleotides can also be used to screen for  
 CC peptide analogues and antagonists.  
 XX  
 SQ Sequence 891 BP; 205 A; 188 C; 170 G; 191 T; 137 other;  
 Query Match 66.4%; Score 16.6; DB 20; Length 891;  
 Best Local Similarity 79.2%; Pred. No. 1.5e+02;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 tcgtcagggtcagggtcagcagcgtt 24  
 ||||| || || ||||| |||||  
 Db 169 TCCTCAGCAGCANGCCAGCAGGTT 146  
 RESULT 22  
 AAZ16835/C  
 ID AAZ16835 standard; cDNA; 891 BP.  
 XX

AAZ16835;  
 12-OCT-1999 (first entry)  
 Human gene expression product cDNA sequence SEQ ID NO:4305.  
 Human; gene; gene expression product; diagnosis; therapy; probe;  
 detection; mapping; tissue typing; profiling; forensic; cancer;  
 genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.  
 Homo sapiens.  
 WO9938972-A2.  
 05-AUG-1999.  
 28-JAN-1999; 99WO-US01619.  
 03-APR-1998; 98US-0080666.  
 28-JAN-1998; 98US-0072910.  
 24-FEB-1998; 98US-0075954.  
 31-MAR-1998; 98US-0080114.  
 03-APR-1998; 98US-0080515.  
 (CHIR) CHIRON CORP.  
 (HYSE-) HYSEQ INC.  
 Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;  
 Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;  
 Jones WL, Kassam A, Kennedy GC, Kita D, Labat J;  
 Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;  
 Stache-Crain B, Sudduth-Klinger J, Williams LT;  
 WPI: 1999-494092/41.  
 Novel human genes and their expression products which are  
 differentially expressed in different cell types  
 Claim 1; Page 2039; 2479pp; English.  
 The present invention describes a library of human polynucleotides  
 comprising the sequences given in AAZ12532 to AAZ1779. Also described is  
 a method of detecting differentially expressed genes correlated with the  
 cancerous state of a mammalian cell, comprising detecting at least one  
 differentially expressed gene product in a test sample from a cell  
 suspected of being cancerous, where the gene product is encoded by one  
 of the 5248 polynucleotide sequences given in AAZ12532 to AAZ1779. The  
 polynucleotides can be used as a source of primers and probes, which can  
 be used for a variety of purpose, e.g. detection of expression levels,  
 mapping, tissue typing or profiling, forensics, genetic analysis and  
 detection of polymorphisms. Polypeptides encoded by the polynucleotides  
 can be used for raising antibodies for experimental, diagnostic and  
 therapeutic purposes. The polynucleotides may also be used to construct  
 arrays for diagnostics (which may be used to determine function of an  
 encoded protein); and to detect differences in expression levels between  
 two cells (e.g. to identify abnormal or diseased tissue in a human, to  
 identify a genetic predisposition or susceptibility to a disease such as  
 cancer). The polynucleotides of the invention are especially used in the  
 diagnosis, prognosis and management of colorectal cancer, breast cancer,  
 and lung cancer. The polynucleotides can also be used to screen for  
 peptide analogues and antagonists.  
 Query Match 56.4%; Score 16.6; DB 20; Length 891;  
 Best Local Similarity 79.2%; Pred. No. 1.5e+02;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Qy 1 tcgtcaggtgcaggtcagcagctt 24  
 ||||| ||| ||||| |||  
 Db 169 TCGTCAGCAGCAGCCAGCAGGTT 146

RESULT 23  
 AAC77412/C  
 ID AAC77412 standard; cDNA; 1953 BP.  
 XX  
 AAC77412;  
 AC  
 08-FEB-2001 (first entry)  
 DT  
 XX  
 Human ORFX ORF2967 polynucleotide sequence SEQ ID NO:5933.  
 DE  
 XX  
 Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.  
 XX  
 Homo sapiens.  
 OS  
 XX  
 WO200058473-A2.  
 PN  
 XX  
 05-OCT-2000.  
 PD  
 XX  
 31-MAR-2000; 2000WO-US08621.  
 PF  
 XX  
 31-MAR-1999; 99US-0127607.  
 PR  
 XX  
 02-APR-1998; 99US-0127636.  
 PR  
 XX  
 05-APR-1999; 99US-0127728.  
 PR  
 XX  
 30-MAR-2000; 2000US-0540763.  
 XX  
 (CURA-) CURAGEN CORP.  
 PA  
 XX  
 Shimkets RA, Leach M;  
 PI  
 XX  
 WPI: 2000-602362/57.  
 DR  
 XX  
 P-PSDB; AAB43203.  
 DR  
 XX  
 Novel nucleic acids and peptides derived from open reading frame X.  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 Claim 5; Page 5116-5117; 5507pp; English.  
 XX  
 AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 CC antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; antirheumatic; antithyroid;  
 CC antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX  
SQ Sequence 1953 BP; 442 A; 635 C; 554 G; 322 T; 0 other;

Query Match 66.4%; Score 16.6; DB 21; Length 1953;  
Best Local Similarity 82.6%; Pred. No. 1.6e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cgtcagggtcaggtcagcagctt 24  
| | | | | | | | | | | | | | | | | |  
Db 731 CGGACGGTGGGTGTCAGCCGCT 709

RESULT 24  
AAV52499/c  
ID AAV52499 standard; DNA; 735 BP.  
XX  
AC AAV52499;  
XX  
DT 23-OCT-1998 (first entry)  
XX  
DE Streptococcus pneumoniae genome fragment SEQ ID NO:366.  
XX  
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
KW computer readable medium; vaccine; pharmaceutical composition; ds.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO9818931-A2.  
XX  
PD 07-MAY-1998.  
XX  
PF 30-OCT-1997; 97WO-US19588.  
XX  
PR 31-OCT-1996; 96US-0029960.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;  
PI Kunsch CA, Rosen CA;  
XX  
DR WPI; 1998-27225/24.  
XX  
PT Computer-readable medium with recorded Streptococcus pneumoniae  
PT polynucleotide sequences - useful in diagnostic kits and assays, and  
PT pharmaceutical compositions and vaccines for Streptococcus  
PT pneumoniae  
XX  
PS Claim 1; Page 1381-1382; 1409pp; English.  
XX  
CC The present invention describes a computer readable medium which has  
CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)  
CC recorded on it, or a representative fragment or a sequence at least 95%  
CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in  
CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from  
CC Streptococcus pneumoniae. The present invention also describes an  
CC isolated nucleic acid molecule encoding a homologue of any of the  
CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the  
CC nucleic acid molecule is produced by a process comprising: (a) screening  
CC a genomic DNA library using as a probe a target sequence defined by any  
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the  
CC library which contain sequences that hybridise to the target sequence and  
CC isolating the nucleic acid molecules from the members; or (b) isolating  
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid  
CC molecules whose nucleotide sequence is homologous to amplification  
CC primers derived from the fragment of the S. pneumoniae genome to prime  
CC the amplification and isolating the amplified sequences. The computer  
CC readable medium can be used in a computer-based system for identifying  
CC fragments of the S. pneumoniae genome of commercial importance, or  
CC expression modulating fragments of the S. pneumoniae genome. Products  
CC from the present invention can be used in diagnosis kits and assays, and  
CC pharmaceutical compositions and vaccines for S. pneumoniae.  
XX

SQ Sequence 735 BP; 177 A; 191 C; 146 G; 220 T; 1 other;

Query Match 64.8%; Score 16.2; DB 19; Length 735;  
Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 caggtgcaggtcagtcagcagcttg 25  
| | | | | | | | | | | | | | | | | |  
Db 31 CGGGTGTGTCAGCAGCGTG 11

RESULT 25  
AAX30776  
ID AAX30776 standard; DNA; 1299 BP.  
XX  
AC AAX30776;  
XX  
DT 20-MAY-1999 (first entry)  
XX  
DE Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:53.  
XX  
KW Streptococcus pneumoniae strain 0100993; vaccine; immune response;  
KW streptococcal infection; pneumococcal; ss.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO9737026-A1.  
XX  
PD 09-OCT-1997.  
XX  
PF 01-APR-1997; 97WO-US05306.  
XX  
PR 22-AUG-1996; 96US-0025788.  
PR 02-APR-1996; 96US-0014690.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
XX  
PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;  
PI Stodola RK;  
XX  
DR WPI; 1997-503111/46.  
XX  
PT P-PSDB; AAY11177.  
XX  
PT Nucleic acids encoding pneumococcal polypeptide(s) - useful in  
PT vaccines, drug screening, etc  
XX  
PS Claim 5; Page 88-89; 354pp; English.  
XX  
CC AAX30724 to AAX30946 represent genomic DNA sequences isolated from  
CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences  
CC encode the novel proteins given in AAY11114 to AAY11367. The proteins,  
CC isolated from Streptococcus pneumoniae, can be used in vaccines against  
CC streptococcal infections and in assays for identifying compounds that  
CC inhibit or activate the activity of the proteins. The antagonists can  
CC be used to treat an individual having need to inhibit a bacterial  
CC protein. Vectors expressing the proteins can be used to induce a  
CC protective immune response in mammals.  
XX  
SQ Sequence 1299 BP; 377 A; 261 C; 341 G; 320 T; 0 other;

Query Match 64.8%; Score 16.2; DB 18; Length 1299;  
Best Local Similarity 85.7%; Pred. No. 2.3e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 caggtgcaggtcagtcagcagcttg 25  
| | | | | | | | | | | | | | | | | |  
Db 1030 cgggtgctggtcagcagcggcg 1050

RESULT 26

AAN71388  
 ID AAN71388 standard; DNA; 1496 BP.  
 XX  
 AC AAN71388;  
 XX  
 DT 02-MAY-1991 (first entry)  
 DE Encodes subtilisin having mutation at amino acid position 222.  
 XX  
 KW subtilisin; extracellular protease; amylase production; ss.  
 XX  
 OS Bacillus amyloliquefaciens.  
 XX  
 FH Key Location/Qualifiers  
 FT mat\_peptide 417..1241  
 FT /\*tag= a  
 FT /product= subtilisin  
 FT sig\_peptide 96..191  
 FT /\*tag= b  
 FT /note= "includes pre-sequence (111-191)"  
 FT RBS 83..88  
 FT /\*tag= c  
 FT terminator 1248..1289  
 FT /\*tag= d  
 FT -10\_signal 23..28  
 FT /\*tag= e  
 FT -35\_signal 1..6  
 FT /\*tag= f  
 XX  
 PN EP246678-A.  
 XX  
 PD 25-NOV-1987.  
 XX  
 PF 01-JAN-1987; 87EP-0200690.  
 XX  
 PR 01-JAN-1987; 87EP-0200690.  
 PR 24-JUN-1983; 83US-0507419.  
 PR 29-MAY-1984; 84US-0614491.  
 PR 29-MAY-1984; 84US-0614612.  
 PR 29-MAY-1984; 84US-0614615.  
 PR 29-MAY-1984; 84US-0614616.  
 PR 29-MAY-1984; 84US-0614617.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Bott RR, Ferrari E, Wells JA, Estell DA, Henner DJ;  
 XX  
 DR WPI: 1987-328920/47.  
 DR P-PSDB; AAP71731.  
 XX  
 PT Bacillus strains not excreting subtilisin or neutral protease -  
 PT obtd. by recombinant DNA procedures, useful for enzyme prodn.  
 PT esp. of hydrolase(s) such as amylase  
 XX  
 PS Example 13; Fig 12; 71pp; English.  
 XX  
 CC The mutation was engineered using standard techniques. When the  
 CC wild-type Met(222) residue (the numbering refers to the mature  
 CC enzyme) is substituted by Ala, the mutant enzyme is considerably  
 CC more stable to oxidation than the wild-type subtilisin. Substituting  
 CC Ser at position 222 also gives improved oxidation stability. Such  
 CC mutants can be used with surfactants or detergents in washing  
 CC compositions.  
 XX  
 SQ Sequence 1496 BP; .421 A; 361 C; 354 G; 357 T; 3 other;

Query Match 64.88; Score 16.2; DB 8; Length 1496;  
 Best Local Similarity 75.08; Pred. No. 2.4e+02;  
 Matches 16; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 cgtcagggtcaggtcagcaggtg 25

Db 1075 cgtcannngcatctccgacgttg 1098  
 RESULT 27  
 AAZ28269/c  
 ID AAZ28269 standard; cDNA; 2517 BP.  
 XX  
 AC AAZ28269;  
 XX  
 DT 05-JAN-2000 (first entry)  
 DE Rat neuronal immediate early gene cDNA clone L048.  
 XX  
 KW Immediate early gene; IEG; neuron; brain; function; growth factor;  
 KW transcription factor; signal transduction; cytoskeletal protein;  
 KW metabolic enzyme; learning; memory; synaptic transmission; tolerance;  
 KW neuronal plasticity; cysteine rich; zinc finger; ds.  
 XX  
 OS Rattus sp.  
 FH Key Location/Qualifiers  
 FT CDS 414..1055  
 FT /\*tag= a  
 FT /product= "L024 protein"  
 XX  
 PN W09940225-A1.  
 XX  
 PD 12-AUG-1999.  
 XX  
 PF 05-FEB-1999; 99WO-US02462.  
 XX  
 PR 09-FEB-1998; 98US-0074135.  
 PR 12-FEB-1998; 98US-0074518.  
 XX  
 PA (UJJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
 PA (BADI ) BASF-LYNX BIOSCIENCE AG.  
 XX  
 PI Worley PF, Lanahan A, Goetz B, Hiemisch H, Kuner R, Scheek S;  
 PI Nikolic K, Zhukovski E;  
 XX  
 DR WPI: 1999-590697/50.  
 DR P-PSDB; AAY42772.  
 XX  
 PT Novel genes and polypeptides, useful for treating conditions related to  
 PT a deficiency in nIEG responsiveness to a stimulus  
 XX  
 PS Claim 1; Page 56-58; 134pp; English.  
 XX  
 CC This sequence represents rat neuronal immediate early gene (IEG) cDNA  
 CC clone L024. This clone encodes a protein which is cysteine rich  
 CC and has a motif with distant homology to that of proteins  
 CC which contain zinc fingers. An IEG is a gene whose expression  
 CC is rapidly increased immediately following a stimulus e.g.,  
 CC neuronal stimulation. Such neuronal IEGs have been found to encode a  
 CC variety of proteins, including transcription factors, cytoskeletal  
 CC proteins, growth factors and metabolic enzymes, as well as proteins  
 CC involved in signal transduction. The identification of neuronal IEGs and  
 CC the proteins they encode may provide important information about the  
 CC function of neurons in, for example, learning, memory, synaptic  
 CC transmission, tolerance and neuronal plasticity. Neuronal IEGs, neuronal  
 CC IEG protein products, cells expressing neuronal IEGs and antibodies  
 CC against neuronal IEG proteins can be used to treat an animal with a  
 CC deficiency in neuronal IEG responsiveness to stimuli, such that the effect  
 CC of the deficiency is minimised. The deficiency may be a reduced or  
 CC elevated level of expression of an IEG. The neuronal stimulus comprises a  
 CC maximal electroconvulsive seizure and its effects influence learning or  
 CC memory. The IEGs and protein products are useful in identifying compounds  
 CC that modulate the expression or activity of IEG nucleic acids or  
 CC proteins, respectively.  
 XX  
 SQ Sequence 2517 BP; 699 A; 541 C; 569 G; 708 T; 0 other;





KW computer readable medium; vaccine; pharmaceutical composition; ds.  
 XX Streptococcus pneumoniae.  
 OS W09818931-A2.  
 PN 07-MAY-1998.  
 PD 30-OCT-1997; 97WO-US19588.  
 XX 31-OCT-1996; 96US-0029960.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;  
 PI Kunsch CA, Rosen CA;  
 XX WPI; 1998-272225/24.  
 DR Computer-readable medium with recorded Streptococcus pneumoniae  
 XX polynucleotide sequences - useful in diagnostic kits and assays, and  
 PT pharmaceutical compositions and vaccines for Streptococcus  
 PT pneumoniae  
 PS Claim 1; Page 1151-1157; 1409pp; English.  
 XX The present invention describes a computer readable medium which has  
 CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)  
 CC recorded on it, or a representative fragment or a sequence at least 95%  
 CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in  
 CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from  
 CC Streptococcus pneumoniae. The present invention also describes an  
 CC isolated nucleic acid molecule encoding a homologue of any of the  
 CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the  
 CC nucleic acid molecule is produced by a process comprising: (a) screening  
 CC a genomic DNA library using as a probe a target sequence defined by any  
 CC of the sequences in SEQ ID NO:1 to 391, identifying members of the  
 CC library which contain sequences that hybridise to the target sequence and  
 CC isolating the nucleic acid molecules from the members; or (b) isolating  
 CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid  
 CC molecules whose nucleotide sequence is homologous to amplification  
 CC primers derived from the fragment of the S. pneumoniae genome to prime  
 CC the amplification and isolating the amplified sequences. The computer  
 CC readable medium can be used in a computer-based system for identifying  
 CC fragments of the S. pneumoniae genome of commercial importance, or  
 CC expression modulating fragments of the S. pneumoniae genome. Products  
 CC from the present invention can be used in diagnosis kits and assays, and  
 CC pharmaceutical compositions and vaccines for S. pneumoniae.  
 XX Sequence 10357 BP; 2979 A; 2388 C; 1821 G; 3169 T; 0 other;

Query Match 64.8%; Score 16.2; DB 19; Length 10357;  
 Best Local Similarity 85.7%; Pred. No. 2.9e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 cagggtcagggtcagcacttg 25  
 | ||||| ||||| ||||| ||  
 Db 2545 CGGGTGTGGTACGACGCGTG 2525

RESULT 31  
 AAV35000  
 ID AAV35000 standard; DNA; 49272 BP.

XX AAV35000;

XX 27-AUG-1998 (first entry)

XX Mycobacteriophage D29 DNA.

XX Shuttle plasmid; vector; mycobacteriophage; lambda phage cosmid; antigen;  
 KW anti-tumour agent; lymphokine; immune response; tolerance; mutant; toxin;

KW autoimmune disease; antisense; vaccine; humoral response; T-cell memory;  
 KW cellular response; mucosal response; diagnostic; antibody;  
 XX infectious agent; ds.  
 OS Mycobacteriophage D29.  
 XX US5773267-A.  
 PN 30-JUN-1998.  
 PD 07-MAR-1996; 96US-0614770.  
 XX 07-MAR-1996; 96US-0614770.  
 PR 07-FEB-1992; 92US-0833431.  
 PR 29-APR-1993; 93US-0057531.  
 PR 23-MAY-1994; 94US-0247901.  
 XX (UYPI-) UNIV PITTSBURGH.  
 PA (YESH ) UNIV YESHIVA EINSTEIN COLLEGE.  
 XX Hatfull GF, Jacobs WR;  
 PI WPI; 1998-387007/33.  
 DR Nucleic acid encoding the D29 mycobacteriophage and related shuttle  
 XX plasmids - for producing recombinant mycobacteria and auxotrophic  
 PT mutants to express antigens and proteins providing long lasting T  
 PT cell memory  
 PS Claim 1; Column 7-50; 28pp; English.

XX This sequence encodes the D29 mycobacteriophage which is used in the  
 CC construction of a new shuttle plasmid containing D29 with an E. coli  
 CC lambda phage cosmid inserted in a non-essential region. This plasmid is  
 CC used to insert nucleic acid into mycobacteria by (non-)homologous or  
 CC site-specific recombination, particularly for expressing antigens,  
 CC anti-tumour agents, lymphokines etc., e.g. for evoking an immune response  
 CC or to induce tolerance to autoimmune diseases such as rheumatoid  
 CC arthritis. Such plasmids that contain a transposon or recombinant DNA can  
 CC be used to generate libraries of mutant mycobacteria, e.g. where  
 CC expression of recombinant mycobacterial DNA (to produce an antisense  
 CC molecule) will inactivate a selected mycobacterial gene. These mutants  
 CC are useful for selection of vaccine strains for inducing humoral,  
 CC cellular and/or mucosal responses and for expressing antigens for use  
 CC as vaccines or diagnostic agents, also other useful proteins such as  
 CC enzymes. Also shuttle plasmids specific for a particular Mycobacterium  
 CC can be used diagnostically, i.e. they are ligated to a reporter gene,  
 CC under control of a strong mycobacterial promoter, then incubated with  
 CC test sample and then examining for expression of the reporter. Vaccines  
 CC based on the new mutants prime long-lasting T cell memory, stimulating  
 CC a secondary antibody response that neutralises infectious agents or  
 CC toxins.  
 XX Sequence 49272 BP; 8912 A; 15900 C; 15401 G; 9059 T; 0 other;

Query Match 64.8%; Score 16.2; DB 19; Length 49272;  
 Best Local Similarity 85.7%; Pred. No. 3.4e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 tcgtcagggtcagggtcagcac 21  
 || ||||| ||||| |||||  
 Db 30020 tcatcagggtgaggtccgcac 30040

RESULT 32  
 AAZ46040/C  
 ID AAZ46040 standard; cDNA; 258 BP.  
 XX AC AAZ46040;

XX 25-APR-2000 (first entry)  
 XX

Partial cDNA encoding a transferase designated HUTRAN-3 fragment.

Transferase; HUTRAN-3; arginine methyltransferase; autoimmune disorder; inflammatory disorder; AIDS; atherosclerosis; adult respiratory distress syndrome; allergy; asthma; trauma; autoimmune thyroiditis; bronchitis; Crohn's disease; diabetes mellitus; gout; Grave's disease; osteoarthritis; osteoporosis; pancreatitis; psoriasis; rheumatoid arthritis; infection; neurological disorder; epilepsy; ischemic cerebrovascular disease; stroke; Alzheimer's disease; Picks disease; Huntington's disease; dementia; Parkinson's disease; extrapyramidal disorder; viral central nervous system disease; prion disease; central nervous system developmental disorder; neuroskeletal disorder; muscular dystrophy; neuromuscular disorder; peripheral nervous system disorder; mental disorder; schizophrenia; anxiety; reproductive disorder; gastrointestinal disorder; dyspepsia; indigestion; gastritis; anorexia; nausea; abdominal angina; gastroenteritis; intestinal obstruction; peptic ulcer; irritable bowel syndrome; diarrhoea; constipation; gastrointestinal haemorrhage; cancer; ss.

Homo sapiens.

OS  
XX  
XX  
PN WO200000594-A2.  
XX  
XX  
PD 06-JAN-2000.  
XX  
XX  
PF 29-JUN-1999; 99WO-US14651.  
XX  
XX  
PR 30-JUN-1998; 98US-0109204.  
XX  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
XX  
PI Lal P, Bandman O, Hillman JL, Guegler KJ, Gorgone GA, Corley NC;  
PI Patterson C;  
XX  
XX  
DR WPI; 2000-147267/13.  
XX  
XX  
PT Novel human transferases used for the diagnosis, treatment, and  
PT prevention of autoimmune/inflammatory, neurological, reproductive and  
PT gastrointestinal disorders and cancer -  
XX  
XX  
PS Disclosure; Page 89; 95pp; English.

AA246039-47 represent partial fragments which were used to produce  
CC Incyte clone 2525071, which encodes a transferase polypeptide,  
CC designated HUTRAN-3. The polypeptide is an arginine methyltransferase.  
CC The HUTRAN polypeptides, polynucleotides, agonists, antagonists, and  
CC antibodies can be used to diagnose, treat or prevent autoimmune/  
CC inflammatory diseases (e.g. AIDS, adult respiratory distress syndrome,  
CC allergies, asthma, atherosclerosis, autoimmune thyroiditis, bronchitis,  
CC Crohn's disease, diabetes mellitus, gout, Grave's disease,  
CC osteoarthritis, osteoporosis, pancreatitis, psoriasis, rheumatoid  
CC arthritis, infections, trauma, neurological disorders (e.g. epilepsy,  
CC ischemic cerebrovascular disease, stroke, Alzheimer's disease, Picks  
CC disease, Huntington's disease, dementia, Parkinson's disease and other  
CC extrapyramidal disorders, viral central nervous system disease, prion  
CC diseases, central nervous system developmental disorders, neuroskeletal  
CC disorders, muscular dystrophy, neuromuscular disorders, peripheral  
CC nervous system disorders, mental disorders, schizophrenia, anxiety,  
CC reproductive disorders and gastrointestinal disorders (e.g. dyspepsia,  
CC indigestion, gastritis, anorexia, nausea, abdominal angina,  
CC gastroenteritis, intestinal obstruction, intestinal tract infections,  
CC peptic ulcer, irritable bowel syndrome, diarrhoea, constipation,  
CC gastrointestinal haemorrhage, and cancer.

XX  
SQ Sequence 258 BP; 55 A; 82 C; 65 G; 55 T; 1 other;

Query Match 64.0%; Score 16; DB 21; Length 258;  
Best Local Similarity 76.0%; Pred. No. 2.4e+02;  
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 1 tcgtcagggtgcaggtcagcagcttg 25

Db 257 TGGTAAGTGCNTGTCCGACGTAG 233  
RESULT 33  
AAFL12855/c  
ID AAF12855 standard; cDNA; 403 BP.  
XX  
XX AAF12855;  
XX  
DT 13-MAR-2001 (first entry)  
XX  
XX Aspergillus oryzae EST SEQ ID NO:5378.  
XX  
XX Multiple gene expression; filamentous fungal cell; EST;  
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
KW culture condition; environmental stress; spore morphogenesis;  
KW metabolic pathway engineering; catabolic pathway engineering; ss.  
XX  
XX Aspergillus oryzae.  
OS  
XX  
XX WO200056762-A2.  
XX  
XX 28-SEP-2000.  
XX  
XX 22-MAR-2000; 2000WO-US07781.  
PF  
XX  
XX 22-MAR-1999; 99US-0273623.  
PR  
XX  
XX (NOVO ) NOVO NORDISK BIOTECH INC.  
PA (NOVO ) NOVO NORDISK AS.  
XX  
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
PI WPI; 2000-594572/56.  
XX  
XX  
XX Monitoring differential expression of genes in filamentous fungal cells  
PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
PT substrate of expressed sequence tags -  
XX  
XX  
XX Claim 88; Page 2240; 3161pp; English.

The present invention describes a method for monitoring differential  
CC expression of genes in a first filamentous fungal (FF) cell relative to  
CC expression of the same genes in one or more second filamentous fungal  
CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
CC are used in the methods for monitoring differential expression of genes  
CC in a first filamentous fungal (FF) cell relative to expression of the  
CC same genes in one or more second filamentous fungal cells. Monitoring  
CC the global expression of genes from FF cells allows the production  
CC potential of the microorganisms to be improved. New genes may be  
CC discovered, possible functions of unknown open reading frames can be  
CC identified and gene copy number variation and stability can be  
CC monitored. The expression of genes can be used to study how FF cells  
CC adapt to changes in culture conditions, environmental stress, spore  
CC morphogenesis, recombination, metabolic or catabolic pathway  
CC engineering. Using ESTs provides several advantages over genomic or  
CC random cDNA clones including elimination of redundancy as one spot on an  
CC array equals one gene or open reading frame, and organisation of the  
CC microarrays based on function of the gene products to facilitate  
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from  
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus  
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and  
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
CC all specifically claimed in the present invention.

XX  
SQ Sequence 403 BP; 94 A; 104 C; 106 G; 97 T; 2 other;

Query Match 64.0%; Score 16; DB 21; Length 403;  
Best Local Similarity 79.2%; Pred. No. 2.5e+02;

```
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Oy 1 tcgtcaggtgcaggtcagcagtt 24
    ||||| ||||| ||||| ||
Db 292 tcattcaggtgcaggtcagcagtt 269
RESULT 34
AAC43677/c
ID AAC43677 standard; DNA; 1188 BP.
XX AC AAC43677;
XX DT 18-OCT-2000 (first entry)
XX ZE Zea mays DNA fragment SEQ ID NO: 40096.
DE XX
DE XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic;
KW pathway; promoter; termination sequence; corn; ss.
XX OS Zea mays subsp. mays.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
```



XX Leucyl tRNA synthetase; leus polypeptide; vaccine;  
 KW genetic immunisation; antibacterial; antibiotic; otitis media;  
 KW conjunctivitis; pneumonia; bacteraemia; meningitis; sinusitis;  
 KW pleural emphysema; endocarditis; gene therapy; ss.  
 XX Streptococcus pneumoniae.

XX Key Location/Qualifiers  
 FH CDS 1..2502  
 FT /\*\*tag= a  
 FT /product= "leus"

XX W09739022-A1.

XX 23-OCT-1997.

XX 18-APR-1997; 97WO-US06875.

XX 18-APR-1996; 96GB-0007993.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

XX (SMIK ) SMITHKLINE BEECHAM PLC.

XX Lawlor EJ;

XX WPI; 1997-526396/48.

XX P-PSDB; AAW32114.

XX Streptococcus pneumoniae leucyl tRNA synthetase - useful to produce  
 PT antibodies or to screen for (ant)agonists with antibacterial  
 PT activity, e.g. to diagnose and treat meningitis, pneumonia, etc.

XX Claim 4; Page 35-36; 48pp; English.

XX The present sequence encodes leucyl tRNA synthetase (leus) from  
 CC Streptococcus pneumoniae. The leus polypeptides, antagonists, antibodies  
 CC and related nucleic acids can be used for diagnosis and treatment of  
 CC bacterial diseases. In particular, they are directed towards  
 CC Streptococcus pneumoniae infections causing otitis media,  
 CC conjunctivitis, pneumonia, bacteraemia, meningitis, sinusitis, pleural  
 CC emphysema and endocarditis. Leus polypeptides, or vectors for their  
 CC expression, can be used prophylactically in vaccines to raise an  
 CC antibody and/or T cell immune response against these same diseases.  
 CC Additionally, the new polypeptides allow agonists and antagonists of  
 CC leus to be identified using standard binding assays. The compounds which  
 CC are identified may have useful bacteriostatic and/or bacteriocidal  
 CC activity.

XX Sequence 2502 BP; 744 A; 524 C; 586 G; 648 T; 0 other;

Query Match 64.0%; Score 16; DB 18; Length 2502;  
 Best Local Similarity 79.2%; Pred. No. 3.1e+02;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cgtcaggtgcaggtcagcagcttg 25  
 |||||  
 Db 128 cgtctggagctggtctgcagctag 151

RESULT 37

AAA79731/C

ID AAA79731 standard; cDNA; 2586 BP.

XX AAA79731;

XX 27-NOV-2000 (first entry)

XX Eucalyptus grandis cell signalling involved polynucleotide SEQ ID NO:883.

XX Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;  
 KW plant cell signalling; modulation; transgenic plant; pathogen; growth;

XX environmental change; development; cell proliferation; differentiation;  
 KW elongation; survival; disease resistance; nutrient metabolism; ss.  
 XX Eucalyptus grandis.

XX W0200042171-A1.

XX 20-JUL-2000.

XX 11-JAN-2000; 2000WO-US00724.

XX 12-JAN-1999; 99US-0228986.

XX 01-NOV-1999; 99US-0162866.

XX (GENE-) GENESIS RES & DEV CORP LTD.

XX Strabala TJ, Nieuwenhuizen NJ;

XX WPI; 2000-476052/41.

XX Isolated polynucleotide encoding a polypeptide involved in cell  
 PT signaling used for generating transgenic plants with modified responses  
 PT to external signals

XX Claim 1; Page 436-437; 527pp; English.

XX AAA79263 to AAA79736 and AAB25100 to AAB25570 represent polynucleotide  
 CC and protein sequences isolated from eucalyptus (Eucalyptus grandis) or  
 CC pine (Pinus radiata also known as Monterey pine). The protein sequences  
 CC are involved in cell signalling. The polynucleotide and protein  
 CC sequences can be used to modify the response of plant cells to external  
 CC signals e.g. environmental changes or pathogens during the growth and  
 CC development of a plant. They can be used to modify cell proliferation,  
 CC differentiation, elongation and survival, resistance to disease and  
 CC nutrient metabolism. Examples of modifications which can be produced are  
 CC altered fruit ripening and senescence of leaves and flowers e.g. to  
 CC delay senescence and prolong the life of cut flowers or enhance  
 CC senescence of reproductive organs to engineer sterile plants. Other  
 CC modifications can be used to delay senescence in selected cell types or  
 CC organs providing fruit and vegetables which have a longer shelf life  
 CC between harvest and consumption, or to decrease branching frequency in  
 CC forest tree species giving long stretches of valuable knot-free clear  
 CC wood which can be used in solid timber furniture and veneers.

XX Sequence 2586 BP; 547 A; 752 C; 705 G; 582 T; 0 other;

Query Match 64.0%; Score 16; DB 21; Length 2586;  
 Best Local Similarity 79.2%; Pred. No. 3.1e+02;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cgtcaggtgcaggtcagcagcttg 25  
 |||||  
 Db 1014 CGGCACGGCGCGGACGACGCTTG 991

RESULT 38

AAA14818/C

ID AAA14818 standard; DNA; 4376 BP.

XX AAA14818;

XX 08-AUG-2000 (first entry)

XX DNA encoding a Cl endoglucanase polypeptide.

XX Cl endoglucanase; promoter; mutant; expression-regulating region;  
 KW secretion signal sequence; enzyme; cellulase; xylanase; pectinase;  
 KW lipase; protease; ss.

XX Chrysosporium lucknowense.

XX Key

Location/Qualifiers

```
FT CDS 2509..3899
FT /*tag= a
FT /*note= "contains 2 introns"
FT sig_peptide 2509..2590
FT /*tag= b
FT exon 2509..2950
FT /*tag= c
FT /*number= 1
FT intron 2951..3060
FT /*tag= d
FT /*number= 1
FT exon 3061..3385
FT /*tag= e
FT /*number= 2
FT intron 3386..3478
FT /*tag= f
FT /*number= 2
FT exon 3478..3899
FT /*tag= g
FT /*number= 3
XX
PN WO200020555-A2.
XX
XX 13-APR-2000.
XX
XX 06-OCT-1999; 99WO-NL00618.
XX
XX 06-OCT-1998; 98WO-EP06496.
XX
XX (AARL-) AARL INC.
XX
XX Emalfarb MA, Burlingame RP, Olson PT, Sinitsyn AP, Parriche M;
XX Bousson JC, Pynnonen CM, Punt PJ, Van Zeijl CMD;
XX
XX WPI; 2000-317705/27.
XX P-PSDB; AAY84797.
XX
XX Mutant Chrysosporium strain comprising nucleic acid encoding proteins,
XX particularly industrially important enzymes such as cellulases,
XX xylanases, pectinases, lipases and proteases -
XX
XX Disclosure; Page 43-46; 92pp; English.
XX
XX The present sequence encodes a Chrysosporium Cl endoglucanase. The
XX polynucleotide sequence comprises a promoter sequence, and can be
XX linked to a nucleic acid sequence of the invention to control
XX expression. The specification describes a mutant Chrysosporium
XX strain which comprises a nucleic acid sequence encoding a polypeptide
XX of interest, where the nucleic acid sequence is linked to an
XX expression-regulating region and a secretion signal sequence.
XX The mutant strain is useful for producing heterologous proteins
XX and polypeptides of interest, e.g. industrially useful enzymes such
XX as cellulases, xylanases, pectinases, lipases and proteases.
XX
XX Sequence 4376 BP; 934 A; 1312 C; 1248 G; 882 T; 0 other;
XX
XX
XX Query Match 64.0%; Score 16; DB 21; Length 4376;
XX Best Local Similarity 79.2%; Pred. No. 3.2e+02;
XX Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
XX
XX QY 2 cgtcagggtcagggtcagcagcttg 25
XX ||||| ||||| ||||| ||||| |||||
XX Db 1831 CGTCAGGGGGAGGGCAGCCCTTG 1808
XX
XX RESULT 39
XX AAA92454/c
XX ID AAA92454 standard; cDNA; 5837 BP.
XX
XX AC AAA92454;
XX
XX DT 15-JAN-2001 (first entry)
```

```
XX Murine E-selectin homologous cDNA sequence SEQ ID NO:1 from Fig 2.
XX DE
XX Endocytic C lectin family; E-selectin; type C lectin; identification;
XX KW ss.
XX
XX OS Mus sp.
XX
XX PN US6117977-A.
XX
XX PD 12-SEP-2000.
XX
XX PF 24-APR-1997; 97US-0840062.
XX
XX PR 24-APR-1996; 95US-0052524.
XX
XX PR 24-APR-1996; 95US-0637021.
XX
XX PA (GETH ) GENENTECH INC.
XX
XX PI Wu K, Lasky LA;
XX
XX WPI; 2000-601415/57.
XX P-PSDB; AAB23822.
XX
XX New type C lectin protein having a signal sequence, cysteine-rich
XX domain, fibronectin type II domain, 8 type C lectin domains,
XX transmembrane domain and a short cytoplasmic domain, used as molecular
XX markers and hybridization probes -
XX
XX Example; Fig 2; 72pp; English.
XX
XX The present invention describes an isolated type C lectin polypeptide
XX (1) comprising amino acid residues 37-1393, 37-174, 175-229, 234-360,
XX 381-507, 520-645, 667-809, 824-951, 970-1108, 1110-1243, or 1259-1393
XX of the protein sequence given in AAB23822. The first 2 polypeptides are
XX capable of binding to a carbohydrate residue, and the rest of the
XX polypeptides are useful for producing antibodies capable of binding to
XX these 2 polypeptides. Polynucleotide sequence encoding the polypeptides
XX of the invention are useful in the identification and purification of
XX their native ligands, and as molecular markers of the tissues in which
XX they are expressed. They provide valuable sequence motifs, which can be
XX inserted or substituted into other native members of the endocytic type
XX C lectins, and provide hybridisation probes for searching cDNA and
XX genomic libraries for the coding sequence of other type C lectins.
XX Variants of type C lectins may be used therapeutically as competitive
XX inhibitors of the biological activity of native type C lectins. The
XX present sequence is said to encode a murine E-selectin homologous
XX protein. The present sequence is given in Figure 2 of the present
XX invention and is said to be SEQ ID NO:1 but does not correspond to the
XX nucleotide presented in the sequence listing as SEQ ID NO:1.
XX
XX Sequence 5837 BP; 1244 A; 1697 C; 1725 G; 1171 T; 0 other;
XX
XX
XX Query Match 64.0%; Score 16; DB 21; Length 5837;
XX Best Local Similarity 79.2%; Pred. No. 3.3e+02;
XX Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
XX
XX QY 1 tcgtcagggtcagggtcagcagcgtt 24
XX ||| | ||||| ||||| |||||
XX Db 4968 TCCTAGGCTGCAGGTCAGCAGGTT 4945
XX
XX RESULT 40
XX AAV52305
XX ID AAV52305 standard; DNA; 11770 BP.
XX
XX AC AAV52305;
XX
XX XX 23-OCT-1998 (first entry)
XX
XX DT Streptococcus pneumoniae genome fragment SEQ ID NO:172.
XX
XX
```

KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
 KW computer readable medium; vaccine; pharmaceutical composition; ds.  
 XX Streptococcus pneumoniae.  
 OS WO9818931-A2.  
 PN 07-MAY-1998.  
 XX 30-OCT-1997; 97WO-US19588.  
 XX 31-OCT-1996; 96US-0029960.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;  
 PI Kunsch CA, Rosen CA;  
 XX WPI; 1998-272225/24.  
 DR Computer-readable medium with recorded Streptococcus pneumoniae  
 PT polynucleotide sequences - useful in diagnostic kits and assays, and  
 PT pharmaceutical compositions and vaccines for Streptococcus  
 PT pneumoniae  
 XX Claim 1; Page 1094-1101; 1409pp; English.  
 PS The present invention describes a computer readable medium which has  
 CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)  
 CC recorded on it, or a representative fragment or a sequence at least 95%  
 CC identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in  
 CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from  
 CC Streptococcus pneumoniae. The present invention also describes an  
 CC isolated nucleic acid molecule encoding a homologue of any of the  
 CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the  
 CC nucleic acid molecule is produced by a process comprising: (a) screening  
 CC a genomic DNA library using as a probe a target sequence defined by any  
 CC of the sequences in SEQ ID NO:1 to 391, identifying members of the  
 CC library which contain sequences that hybridize to the target sequence and  
 CC isolating the nucleic acid molecules from the members; or (b) isolating  
 CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid  
 CC molecules whose nucleotide sequence is homologous to amplification  
 CC primers derived from the fragment of the S. pneumoniae genome to prime  
 CC the amplification and isolating the amplified sequences. The computer  
 CC readable medium can be used in a computer-based system for identifying  
 CC fragments of the S. pneumoniae genome of commercial importance, or  
 CC expression modulating fragments of the S. pneumoniae genome. Products  
 CC from the present invention can be used in diagnosis kits and assays, and  
 CC pharmaceutical compositions and vaccines for S. pneumoniae.  
 XX Sequence 11770 BP; 3673 A; 1919 C; 2574 G; 3603 T; 1 other;  
 SQ Query Match 64.0%; Score 16; DB 19; Length 11770;  
 Best Local Similarity 79.2%; Pred. No. 3.6e+02;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Qy 2 cgtcaggtgcaggtcagcagcttg 25  
 |||| || || |||| |||| ||  
 Db 7356 cgtctggagctggctgcagctag 7379  
 RESULT 41  
 AAA81487  
 ID AAA81487 standard; DNA; 50925 BP.  
 XX AAA81487;  
 AC  
 XX 04-DEC-2000 (first entry)  
 DT N. meningitidis partial DNA sequence gnm\_35 SEQ ID NO:35.  
 XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW

KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; MenB; ds.  
 XX Neisseria meningitidis.  
 OS WO200022430-A2.  
 PN 20-APR-2000.  
 XX 08-OCT-1999; 99WO-US23573.  
 XX 09-OCT-1998; 98US-0103794.  
 PR 30-APR-1999; 99US-0132068.  
 XX (CHIR ) CHIRON CORP.  
 PA Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
 PI Rappuoli R, Pizza M;  
 XX WPI; 2000-318079/27.  
 DR Isolated nucleotide sequences of Neisseria meningitidis which can be  
 XX used in the diagnosis and treatment of N. meningitidis infection and  
 PT other Neisserial infections, for example, N.gonorrhoea -  
 PT Claim 7; Page 607-622; 1760pp; English.  
 PS The present invention describes methods of obtaining immunogenic  
 XX proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
 CC represent specifically claimed Neisseria meningitidis genomic DNA  
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to  
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to  
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to  
 CC Neisserial bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against Meningococcus B; against all serotypes;  
 CC and/or against all pathogenic Neisseriae. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.  
 XX Sequence 50925 BP; 12504 A; 14157 C; 12638 G; 11626 T; 0 other;  
 SQ Query Match 64.0%; Score 16; DB 21; Length 50925;  
 Best Local Similarity 79.2%; Pred. No. 4.1e+02;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Qy 2 cgtcaggtgcaggtcagcagcttg 25  
 |||| || || |||| |||| ||  
 Db 20791 catccggcaggtgcagcagcttg 20814  
 RESULT 42  
 AAF21610  
 ID AAF21610 standard; DNA; 349980 BP.  
 XX AAF21610;  
 AC  
 XX 13-MAR-2001 (first entry)  
 DT







PR	18-JUN-1999;	9905-0139458;
PR	18-JUN-1999;	9905-0139459;
PR	18-JUN-1999;	9905-0139460;
PR	18-JUN-1999;	9905-0139461;
PR	18-JUN-1999;	9905-0139462;
PR	18-JUN-1999;	9905-0139463;
PR	18-JUN-1999;	9905-0139750;
PR	18-JUN-1999;	9905-0139751;
PR	21-JUN-1999;	9905-0139817;
PR	21-JUN-1999;	9905-0139819;
PR	23-JUN-1999;	9905-0140353;
PR	23-JUN-1999;	9905-0140354;
PR	24-JUN-1999;	9905-0140695;
PR	28-JUN-1999;	9905-0140823;
PR	29-JUN-1999;	9905-0140991;
PR	30-JUL-1999;	9905-0141287;
PR	01-JUL-1999;	9905-0141842;
PR	01-JUL-1999;	9905-0142154;
PR	02-JUL-1999;	9905-0142055;
PR	06-JUL-1999;	9905-0142390;
PR	08-JUL-1999;	9905-0142803;
PR	09-JUL-1999;	9905-0142920;
PR	13-JUL-1999;	9905-0142977;
PR	13-JUL-1999;	9905-0143542;
PR	14-JUL-1999;	9905-0143624;
PR	15-JUL-1999;	9905-0144005;
PR	16-JUL-1999;	9905-0144085;
PR	16-JUL-1999;	9905-0144086;
PR	19-JUL-1999;	9905-0144325;
PR	19-JUL-1999;	9905-0144334;
PR	19-JUL-1999;	9905-0144335;
PR	20-JUL-1999;	9905-0144352;
PR	20-JUL-1999;	9905-0144632;
PR	20-JUL-1999;	9905-0144884;
PR	21-JUL-1999;	9905-0144814;
PR	21-JUL-1999;	9905-0144833;
PR	19-JUL-1999;	9905-0144332;
PR	19-JUL-1999;	9905-0144334;
PR	19-JUL-1999;	9905-0144335;
PR	20-JUL-1999;	9905-0144352;
PR	20-JUL-1999;	9905-0144632;
PR	20-JUL-1999;	9905-0144884;
PR	21-JUL-1999;	9905-0144814;
PR	21-JUL-1999;	9905-0144833;
PR	22-JUL-1999;	9905-0145085;
PR	22-JUL-1999;	9905-0145085;
PR	22-JUL-1999;	9905-0145087;
PR	22-JUL-1999;	9905-0145089;
PR	22-JUL-1999;	9905-0145192;
PR	23-JUL-1999;	9905-0145145;
PR	23-JUL-1999;	9905-0145218;
PR	23-JUL-1999;	9905-0145224;
PR	26-JUL-1999;	9905-0145276;
PR	27-JUL-1999;	9905-0145913;
PR	27-JUL-1999;	9905-0145918;
PR	27-JUL-1999;	9905-0145919;
PR	28-JUL-1999;	9905-0145951;
PR	02-AUG-1999;	9905-0146386;
PR	02-AUG-1999;	9905-0146388;
PR	02-AUG-1999;	9905-0146389;
PR	03-AUG-1999;	9905-0147038;
PR	04-AUG-1999;	9905-0147204;
PR	04-AUG-1999;	9905-0147302;
PR	05-AUG-1999;	9905-0147192;
PR	06-AUG-1999;	9905-0147303;
PR	06-AUG-1999;	9905-0147416;
PR	09-AUG-1999;	9905-0147493;
PR	09-AUG-1999;	9905-0147935;
PR	10-AUG-1999;	9905-0148319;
PR	11-AUG-1999;	9905-0148319;
PR	12-AUG-1999;	9905-0148341;
PR	13-AUG-1999;	9905-0148684;
PR	16-AUG-1999;	9905-0149368;
PR	17-AUG-1999;	9905-0149175;
PR	18-AUG-1999;	9905-0149426;
PR	20-AUG-1999;	9905-0149722;

Search completed: October 9, 2001, 11:40:31  
Job time: 1946 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:40:04 ; Search time 221.13 seconds  
(without alignments)  
21.403 Million cell updates/sec

Title: US-09-396-196F-10

Perfect score: 25

Sequence: 1 tcgtcaggtgcaggtcagcagcttg 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

- Issued\_Patents\_NA:\*
- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
  - 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
  - 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*
  - 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*
  - 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*
  - 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	17.6	70.4	7286	5	PCT-US95-11684-1
6	17	68.0	2095	1	US-08-336-408B-3
7	17	68.0	2095	5	PCT-US91-00399-3
8	17	68.0	20303	1	US-08-370-975B-6
9	17	68.0	26764	1	US-08-370-975B-1
10	16.2	64.8	48	6	RE34606-11
11	16.2	64.8	49272	1	US-08-614-770A-1
12	16	64.0	258	3	US-09-109-204-22
13	16	64.0	1866	1	US-08-336-408B-1
14	16	64.0	1866	5	PCT-US91-00399-1
15	16	64.0	2502	2	US-08-844-086-1
16	16	64.0	2502	3	US-09-018-211-1
17	15.8	63.2	987	1	US-08-186-529-1
18	15.8	63.2	987	1	US-08-640-386A-1
19	15.8	63.2	987	4	US-08-848-760B-24
20	15.8	63.2	1018	2	US-08-184-009-194
21	15.8	63.2	1018	2	US-08-458-356-194
22	15.8	63.2	1399	2	US-08-751-767A-3
23	15.8	63.2	1560	2	US-08-751-767A-11
24	15.8	63.2	1623	2	US-08-751-767A-9
25	15.8	63.2	2362	1	US-08-265-087-1
26	15.8	63.2	2362	1	US-08-621-493-1
27	15.8	63.2	2362	2	US-08-965-688-1

28	15.8	63.2	2362	4	US-09-260-173-1	Sequence 1, Appli
29	15.8	63.2	2655	1	US-08-471-033-17	Sequence 17, Appli
30	15.8	63.2	2655	1	US-08-471-033-26	Sequence 26, Appli
31	15.8	63.2	2655	2	US-08-471-044-17	Sequence 17, Appli
32	15.8	63.2	2655	2	US-08-471-044-26	Sequence 26, Appli
33	15.8	63.2	2655	2	US-08-463-483A-17	Sequence 17, Appli
34	15.8	63.2	2655	2	US-08-463-483A-26	Sequence 26, Appli
35	15.8	63.2	2655	2	US-08-471-046A-17	Sequence 17, Appli
36	15.8	63.2	2655	2	US-08-471-046A-26	Sequence 26, Appli
37	15.8	63.2	2655	2	US-08-470-566B-17	Sequence 17, Appli
38	15.8	63.2	2655	2	US-08-470-566B-26	Sequence 26, Appli
39	15.8	63.2	2655	2	US-08-469-334-17	Sequence 17, Appli
40	15.8	63.2	2655	2	US-08-469-334-26	Sequence 26, Appli
41	15.8	63.2	2655	3	US-09-300-529-17	Sequence 17, Appli
42	15.8	63.2	2655	3	US-09-300-529-26	Sequence 26, Appli
43	15.8	63.2	6139	2	US-08-751-767A-7	Sequence 7, Appli
44	15.6	62.4	796	4	US-08-927-219-36	Sequence 36, Appli
45	15.6	62.4	907	3	US-08-767-942A-12	Sequence 12, Appli

ALIGNMENTS

RESULT 1  
US-08-401-068-7  
; Sequence 7, Application US/08401068  
; Patent No. 5859335  
; GENERAL INFORMATION:  
; APPLICANT: Patton, David  
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30B  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/401,068  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/351,970  
; FILING DATE: 08-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Elmer, James Scott  
; REGISTRATION NUMBER: 36,129  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8614  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1041 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHEetical: NO  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1038  
; IDENTIFICATION METHOD: experimental  
; OTHER INFORMATION: /product= "biotin synthase"  
; OTHER INFORMATION: /evidence= EXPERIMENTAL  
US-08-401-068-7

Query Match 100.08; Score 25; DB 2; Length 1041;  
 Best Local Similarity 100.08; Pred. No. 0.014;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgtcagggtcaggtcagcacgttg 25  
 |||  
 Db 111 TCGTCAGGTGCAGGTGCAGCACGCTTG 135

## RESULT 2

US-08-846-338-7  
 ; Sequence 7, Application US/08846338  
 ; Patent No. 5869719

## GENERAL INFORMATION:

APPLICANT: Patton, David  
 TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5869719artis Corporation

STREET: 520 White Plains Road, P.O. Box 2005

CITY: Tarrytown

STATE: NY

COUNTRY: USA

ZIP: 10591

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/846.338

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Meigs, J. Timothy

REGISTRATION NUMBER: 38,241

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8587

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1041 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1038

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /product= "biotin synthase"

OTHER INFORMATION: /evidence= EXPERIMENTAL

US-08-846-338-7

Query Match 100.08; Score 25; DB 2; Length 1041;  
 Best Local Similarity 100.08; Pred. No. 0.014;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgtcagggtcaggtcagcacgttg 25  
 |||  
 Db 111 TCGTCAGGTGCAGGTGCAGCACGCTTG 135

## RESULT 3

US-08-411-768B-1

; Sequence 1, Application US/08411768B

; Patent No. 6083712

GENERAL INFORMATION:

APPLICANT: Olwen Birch

APPLICANT: Johann Brass

APPLICANT: Martin Fuhrmann

APPLICANT: Nicholas Shaw  
 TITLE OF INVENTION: Biotechnological Method  
 TITLE OF INVENTION: of Producing Biotin  
 NUMBER OF SEQUENCES: 19  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
 STREET: 30 Rockefeller Plaza  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10112  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: WordPerfect  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/411,768B  
 FILING DATE: 31-March-95  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: CH 3124/92  
 FILING DATE: 02-OCT-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: CH 2134/93  
 FILING DATE: 15-JUL-1993  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5872 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Escherichia coli  
 STRAIN: DSM498  
 IMMEDIATE SOURCE:  
 CLONE: pBO30A-15/9  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 117..1157  
 IDENTIFICATION METHOD: experimental  
 OTHER INFORMATION: /codon\_start= 117  
 OTHER INFORMATION: /product= "Biotin synthase"  
 OTHER INFORMATION: /evidence= EXPERIMENTAL  
 OTHER INFORMATION: /gene= "bioB"  
 OTHER INFORMATION: /number= 1  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 2295..3050  
 OTHER INFORMATION: /codon\_start= 2295  
 OTHER INFORMATION: /function= "involved in pimeloyl-CoA synthesis"  
 OTHER INFORMATION: /product= "protein"  
 OTHER INFORMATION: /gene= "bioC"  
 OTHER INFORMATION: /number= 3  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 3750..5039  
 IDENTIFICATION METHOD: experimental  
 OTHER INFORMATION: /codon\_start= 3750  
 OTHER INFORMATION: /EC\_number= 2.6.1.62  
 OTHER INFORMATION: /product= "DAPA synthase"  
 OTHER INFORMATION: /evidence= EXPERIMENTAL  
 OTHER INFORMATION: /gene= "bioA"  
 OTHER INFORMATION: /number= 5  
 OTHER INFORMATION: /standard\_name=  
 OTHER INFORMATION: "S-Adenosyl-L-methionine:8-amino-7-oxononanoate  
 OTHER INFORMATION: aminotransf."  
 FEATURE:  
 NAME/KEY: CDS

```
; LOCATION: 5098..5574
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 5098
; OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
; OTHER INFORMATION: /product= "protein"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ORFI"
; OTHER INFORMATION: /number= 6
; FEATURE:
; NAME/KEY: -10_signal
; LOCATION: 45..49
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "promoter ptac"
; FEATURE:
; NAME/KEY: -35_signal
; LOCATION: 23..28
; OTHER INFORMATION: /standard_name= "promoter ptac"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 105..119
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "bioB RBS no.9"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 2284..2297
; OTHER INFORMATION: /standard_name= "bioC RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 3742..3752
; OTHER INFORMATION: /standard_name= "bioA RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 5088..5100
; OTHER INFORMATION: /standard_name= "ORFI RBS"
; FEATURE:
; NAME/KEY: terminator
; LOCATION: 5583..5644
; OTHER INFORMATION: /standard_name= "rho-independent
; OTHER INFORMATION: transcriptional terminator"
; FEATURE:
; NAME/KEY: stem_loop
; LOCATION: 5583..5605
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 1..96
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /function= "promoter ptac"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: WO 87/01391 B1
; FILING DATE: 26-AUG-1986
; PUBLICATION DATE: 07-APR-1993
; US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagcagcttg 25
   |||||
DB 227 TCCTCAGGTGCAGGTGCAGCAGCTTG 251

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
```

```
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411,768B
; FILING DATE: 31-March-95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 3124/92
; FILING DATE: 02-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 2134/93
; FILING DATE: 15-JUL-1993
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5872 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: DSM498
; IMMEDIATE SOURCE:
; CLONE: pB030A15-9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1154..2308
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 1154
; OTHER INFORMATION: /EC_number= 2.3.1.47
; OTHER INFORMATION: /product= "KAPA synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "bioF"
; OTHER INFORMATION: /number= 2
; OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3043..3753
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 3043
; OTHER INFORMATION: /EC_number= 6.3.3.3
; OTHER INFORMATION: /product= "DTB synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "bioD"
; OTHER INFORMATION: /number= 4
; OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 1141..1156
; OTHER INFORMATION: /standard_name= "bioF RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 3030..3045
; OTHER INFORMATION: /standard_name= "bioD RBS"
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: WO 87/01391 B1
```

; FILING DATE: 26-AUG-1986  
; PUBLICATION DATE: 07-APR-1993  
US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;  
Best Local Similarity 100.0%; Pred. No. 0.017;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcagcttg 25  
|||||  
Db 227 TCGTCAGGTGCAGGTACGAGCTTG 251

RESULT 5  
PCT-US95-11684-1/c  
; Sequence 1, Application PC/TUS9511684  
; GENERAL INFORMATION:  
; APPLICANT: THE SCRIPPS RESEARCH INSTITUTE  
; TITLE OF INVENTION: CYTOTACTIN DERIVATIVES THAT STIMULATE  
; TITLE OF INVENTION: ATTACHMENT AND NEURITE OUTGROWTH, AND METHODS OF MAKING  
; TITLE OF INVENTION: AND USING SAME  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: The Scripps Research Institute, Office of  
; ADDRESSEE: Patent Counsel  
; STREET: 10666 North Torrey Pines Road, TPC 8  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/11684  
; FILING DATE: 14-SEP-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/308,359  
; FILING DATE: 16-SEP-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Logan, April C.  
; REGISTRATION NUMBER: 33,950  
; REFERENCE/DOCKET NUMBER: BEC0019P  
; TELEPHONE: 619-554-2937  
; TELEFAX: 619-554-6312  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7286 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 55..6654  
; OTHER INFORMATION: /product= "cytotactin"  
PCT-US95-11684-1

Query Match 70.4%; Score 17.6; DB 5; Length 7286;  
Best Local Similarity 83.3%; Pred. No. 27;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 cytcaggtgcaggtcagcagcttg 25  
|||||  
Db 2200 CTTTCAGGTGCAGGTAACTAGCTGG 2177

RESULT 6  
US-08-336-408B-3/c  
; Sequence 3, Application US/08336408B  
; Patent No. 5723329  
; GENERAL INFORMATION:  
; APPLICANT: EVANS, RONALD M.  
; APPLICANT: MANGELSDORF, DAVID J.  
; TITLE OF INVENTION: RETINOID RECEPTOR COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PRETTY, SCHROEDER, BRUEGEMANN & CLARK  
; STREET: 444 SOUTH FLOWER STREET, SUITE 2000  
; CITY: LOS ANGELES  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/336,408B  
; FILING DATE: 08-NOV-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/933,453  
; FILING DATE: 21-AUG-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/US91/00399  
; FILING DATE: 22-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/478,071  
; FILING DATE: 09-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: REITER, STEPHEN E.  
; REGISTRATION NUMBER: 31,192  
; REFERENCE/DOCKET NUMBER: P41 9851  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-546-1995  
; TELEFAX: 619-546-9392  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2095 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 178..1581  
US-08-336-408B-3

Query Match 68.0%; Score 17; DB 1; Length 2095;  
Best Local Similarity 80.0%; Pred. No. 45;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcagcttg 25  
|||||  
Db 708 TCGGCAGGTGTAGGTGAGGTCTTTG 684

RESULT 7  
PCT-US91-00399-3/c  
; Sequence 3, Application PC/TUS9100399  
; GENERAL INFORMATION:  
; APPLICANT: Mangelsdorf, Dr., David J.  
; APPLICANT: Evans Dr., Ronald M.  
; TITLE OF INVENTION: RETINOID RECEPTOR COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 7



;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: FITCH, EVEN, TABIN & FLANNERY  
;; STREET: 135 South LaSalle Street, Suite 900  
;; CITY: Chicago  
;; STATE: Illinois  
;; COUNTRY: USA  
;; ZIP: 60603  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US91/00399  
;; FILING DATE: 19910122  
;; CLASSIFICATION: 435  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 07/478,071  
;; FILING DATE: 09-FEB-1990  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Reiter, Mr., Stephen E.  
;; REGISTRATION NUMBER: 31192  
;; REFERENCE/DOCKET NUMBER: 50852  
;;  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (619) 552-1311  
;; TELEFAX: (619) 552-0095  
;; TELEX: 20 6566 PATLAW CGO  
;;  
;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2095 base pairs  
;; TYPE: NUCLEIC ACID  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; IMMEDIATE SOURCE:  
;; CLONE: RXR MOUSE ALPHA  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: join(178..1581)  
PCT-US91-00399-3

Query Match 68.0%; Score 17; DB 5; Length 2095;  
Best Local Similarity 80.0%; Pred. No. 45;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagcagttg 25  
||| ||||| ||||| |||||  
DB 708 TCGCAGGTGTAGGTGAGGTCTTTG 684

RESULT 8  
US-08-370-975B-6/c  
; Sequence 6, Application US/08370975B  
; Patent No. 5622851  
; GENERAL INFORMATION:  
; APPLICANT: Maley, Frank  
; APPLICANT: Maley, Gladys F.  
; APPLICANT: Weiner, Karen X.B.  
; TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle  
; STREET: Clinton Square, P.O. Box 1051  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/370,975B  
;; FILING DATE: 10-JAN-1995  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Timian, Susan J.  
;; REGISTRATION NUMBER: 34,103  
;; REFERENCE/DOCKET NUMBER: 20894/80  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (716)263-1636  
;; TELEFAX: (716)263-1600  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 20303 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; POSITION IN GENOME:  
;; CHROMOSOME/SEGMENT: 4q35  
US-08-370-975B-6  
Query Match 68.0%; Score 17; DB 1; Length 20303;  
Best Local Similarity 80.0%; Pred. No. 54;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 tcgtcaggtgcaggtcagcagttg 25  
||| ||||| ||||| ||||| |||||  
DB 285 TGGTGTGTCAGGCCAGCAGGTTG 261  
RESULT 9  
US-08-370-975B-1/c  
; Sequence 1, Application US/08370975B  
; Patent No. 5622851  
; GENERAL INFORMATION:  
; APPLICANT: Maley, Frank  
; APPLICANT: Maley, Gladys F.  
; APPLICANT: Weiner, Karen X.B.  
; TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle  
; STREET: Clinton Square, P.O. Box 1051  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/370,975B  
; FILING DATE: 10-JAN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Timian, Susan J.  
; REGISTRATION NUMBER: 34,103  
; REFERENCE/DOCKET NUMBER: 20894/80  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (716)263-1636  
; TELEFAX: (716)263-1600  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 26764 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; POSITION IN GENOME:

CHROMOSOME/SEGMENT: 4q35  
US-08-370-975B-1

Query Match 68.0%; Score 17; DB 1; Length 26764;  
Best Local Similarity 80.0%; Pred. No. 55;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcagcttg 25  
||| ||||| ||||| |||||

Db 2248 TGGTGTGTGTCAGGCCAGCAGGTTG 2224

RESULT 10

RE34606-11

Patent No. RE34,606

APPLICANT: ESTELL, DAVID A.; WELLS, JAMES A.; BOTT,

RICHARD R.

TITLE OF INVENTION: MODIFIED ENZYMES AND METHODS FOR

MAKING SAME

NUMBER OF SEQUENCES: 27

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/556,918

FILING DATE: 20-JUL-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 614,612

FILING DATE: 29-MAY-1984

SEQ ID NO: 11:

LENGTH: 48

RE34606-11

Query Match

Best Local Similarity 64.8%; Score 16.2; DB 6; Length 48;

Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 cgtcaggtgcaggtcagcagcttg 25

||||| ||||| ||||| |||||

Db 14 cgtcannngcatctccgcagcttg 37

RESULT 11

US-08-614-770A-1

Sequence 1, Application US/08614770A

Patent No. 5773267

GENERAL INFORMATION:

APPLICANT: WILLIAM R. JACOBS AND GRAHAM F. HATFULL

TITLE OF INVENTION: D29 SHUTTLE PHASMSIDS AND USES THEREOF

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN

STREET: 90 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: U.S.A.

ZIP: 10016

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE

MEDIUM TYPE: DISKETTE

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/614,770A

FILING DATE: MARCH 7, 1996

ATTORNEY/AGENT INFORMATION:

NAME: ELIZABETH A. BOGOSIAN

REGISTRATION NUMBER: 39,911

REFERENCE/DOCKET NUMBER: 96700/402

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 697-5995

TELEFAX: (212) 286-0854 or 286-0082

TELEX: TWX 710-581-4765

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 49272

TYPE: NUCLEIC ACID

STRANDEDNESS: DOUBLE

TOPOLOGY: LINEAR

MOLECULE TYPE:

DESCRIPTION: OLIGONUCLEOTIDE

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: MYCOBACTERIOPHAGE

INDIVIDUAL ISOLATE: D29

US-08-614-770A-1

Query Match

Best Local Similarity 64.8%; Score 16.2; DB 1; Length 49272;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagcagcac 21

||| ||||| ||||| |||||

Db 30020 TCATCAGGTGGAGGTCCGCAC 30040

RESULT 12

US-09-109-204-22/c

Sequence 22, Application US/09109204

Patent No. 6060250

GENERAL INFORMATION:

APPLICANT: Lal, Preeti

APPLICANT: Bandman, Olga

APPLICANT: Hillman, Jennifer L.

APPLICANT: Guegler, Karl J.

APPLICANT: Gorgone, Gina A.

APPLICANT: Corley, Neil C.

APPLICANT: Patterson, Chandra

TITLE OF INVENTION: HUMAN TRANSFERASES

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/109,204

FILING DATE: HERewith

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Cerrone, Michael C.

REGISTRATION NUMBER: 39,132

REFERENCE/DOCKET NUMBER: PF-0546 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-855-0572

TELEX:

INFORMATION FOR SEQ ID NO: 22:

SEQUENCE CHARACTERISTICS:

LENGTH: 258 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: BLADTUT07  
CLONE: 1889292H1  
US-09-109-204-22

Query Match 64.0%; Score 16; DB 3; Length 258;  
Best Local Similarity 76.0%; Pred. No. 1e+02;  
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 tgcacaggtgcaggtcagcacgttg 25  
||| ||||| ||| ||||| |  
Db 257 TGGTAAGGTGCTNTGCCGACGCTAG 233

## RESULT 13

US-08-336-408B-1/c  
; Sequence 1, Application US/08336408B  
; Patent No. 5723329  
; GENERAL INFORMATION:  
; APPLICANT: EVANS, RONALD M.  
; APPLICANT: MANGELSDORF, DAVID J.  
; TITLE OF INVENTION: RETINOID RECEPTOR COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PRETTY, SCHROEDER, BRUEGGEMANN & CLARK  
; STREET: 444 SOUTHWEST FLOWER STREET, SUITE 2000  
; CITY: LOS ANGELES  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90071

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/336.408B  
; FILING DATE: 08-NOV-1994  
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/933,453  
; FILING DATE: 21-AUG-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/US91/00399  
; FILING DATE: 22-JAN-1991

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/478,071  
; FILING DATE: 09-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: REITER, STEPHEN E.  
; REGISTRATION NUMBER: 31,192  
; REFERENCE/DOCKET NUMBER: P41 9851  
; TELEPHONE: 619-546-1995  
; TELEFAX: 619-546-9392

INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1866 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: both  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 76..1464  
US-08-336-408B-1

Query Match 64.0%; Score 16; DB 1; Length 1866;  
Best Local Similarity 79.2%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 cgtcaggtgcaggtcagcacgttg 25  
||| ||||| ||||| |  
Db 590 CGGACAGGTGTAGTTCAGGTCCTTG 567

## RESULT 14

PCT-US91-00399-1/c  
; Sequence 1, Application PC/TUS9100399  
; GENERAL INFORMATION:  
; APPLICANT: Mangelsdorf, Dr., David J.  
; APPLICANT: Evans Dr., Ronald M.  
; TITLE OF INVENTION: RETINOID RECEPTOR COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FITCH, EVEN, TABIN & FLANNERY  
; STREET: 135 South LaSalle Street, Suite 900  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60603

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US91/00399  
; FILING DATE: 19910122  
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/478,071  
; FILING DATE: 09-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reiter, Mr., Stephen E.  
; REGISTRATION NUMBER: 31192  
; REFERENCE/DOCKET NUMBER: 50852  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 552-1311  
; TELEFAX: (619) 552-0095

INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1866 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; IMMEDIATE SOURCE:  
; CLONE: RXR HUMAN ALPHA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: join(76..1464)

PCT-US91-00399-1

Query Match 64.0%; Score 16; DB 5; Length 1866;  
Best Local Similarity 79.2%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 cgtcaggtgcaggtcagcacgttg 25  
||| ||||| ||||| |  
Db 590 CGGACAGGTGTAGTTCAGGTCCTTG 567

## RESULT 15

US-08-844-086-1  
; Sequence 1, Application US/08844086  
; Patent No. 5866390  
; GENERAL INFORMATION:  
; APPLICANT: Lawlor, Elizabeth  
; TITLE OF INVENTION: No. 5866390e1 Compounds  
; NUMBER OF SEQUENCES: 6

```
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/844,086
; FILING DATE: 18-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 9607993.4
; FILING DATE: 18-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P31457-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-08-844-086-1

Query Match 64.0%; Score 16; DB 2; Length 2502;
Best Local Similarity 79.2%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cgtcaggtgcaggtcagcacgttg 25
   ||||| ||||| ||||| |||||
Db 128 CGTCTGGAGCTGGTCTGCACGTAG 151

RESULT 16
US-09-018-211-1
; Sequence 1, Application US/09018211
; Patent No. 6048716
; GENERAL INFORMATION:
; APPLICANT: Lawlor, Elizabeth
; TITLE OF INVENTION: No. 6048716el Compounds
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/018,211
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/844,086
; FILING DATE: 18-APR-1997
```

```
;
; APPLICATION NUMBER: 9607993.4
; FILING DATE: 18-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P31457-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-09-018-211-1

Query Match 64.0%; Score 16; DB 3; Length 2502;
Best Local Similarity 79.2%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cgtcaggtgcaggtcagcacgttg 25
   ||||| ||||| ||||| |||||
Db 128 CGTCTGGAGCTGGTCTGCACGTAG 151

RESULT 17
US-08-186-529-1
; Sequence 1, Application US/08186529
; Patent No. 5573764
; GENERAL INFORMATION:
; APPLICANT: Sykes, Megan
; APPLICANT: Wolf, Stanley F.
; TITLE OF INVENTION: USE OF INTERLEUKIN-12 TO PREVENT
; TITLE OF INVENTION: GRAFT-VERSUS-HOST DISEASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc., Legal Affairs
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/186,529
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: McDaniel, Patricia A.
; REGISTRATION NUMBER: 33,194
; REFERENCE/DOCKET NUMBER: GI 5225
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-498-8401
; TELEFAX: 617-876-5851
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 987 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL TYPE: Lymphoblast
; CELL LINE: RPMI 8866
```

FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..987  
US-08-186-529-1

Query Match 63.2%; Score 15.8; DB 1; Length 987;  
Best Local Similarity 89.5%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagc 19  
||| ||||| |||||  
Db 747 TCGGCAGGTGGAGGTCAGC 765

RESULT 18  
US-08-640-386A-1  
; Sequence 1, Application US/08640386A  
; Patent No. 5756085  
; GENERAL INFORMATION:  
; APPLICANT: Sykes, Megan  
; APPLICANT: Wolf, Stanley F.  
; TITLE OF INVENTION: USE OF INTERLEUKIN-12 TO PREVENT  
; TITLE OF INVENTION: GRAFT-VERSUS-HOST DISEASE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc., Legal Affairs  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/640.386A  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brown, Scott A.  
; REGISTRATION NUMBER: 32,724  
; REFERENCE/DOCKET NUMBER: GI 5225A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-498-8224  
; TELEFAX: 617-876-5851  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 987 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; CELL TYPE: Lymphoblast  
; CELL LINE: RPMI 8866  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..987  
US-08-640-386A-1

Query Match 63.2%; Score 15.8; DB 1; Length 987;  
Best Local Similarity 89.5%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagc 19  
||| ||||| |||||  
Db 747 TCGGCAGGTGGAGGTCAGC 765

RESULT 19  
US-08-848-760B-24  
; Sequence 24, Application US/08848760B  
; Patent No. 6248721  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Lung-Ji  
; TITLE OF INVENTION: Animal Model For Evaluation Of Vaccines  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: Florida  
; COUNTRY: United States of America  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/848.760B  
; FILING DATE: 25-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/838,702  
; FILING DATE: 09-APR-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PAGE, DORAN R.  
; REGISTRATION NUMBER: 38,261  
; REFERENCE/DOCKET NUMBER: CNG-100C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (352) 375-8100  
; TELEFAX: (352) 372-5800  
; INFORMATION FOR SEQ ID NO: 24:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 987 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "DNA"  
; SEQUENCE DESCRIPTION: SEQ ID NO: 24:  
US-08-848-760B-24

Query Match 63.2%; Score 15.8; DB 4; Length 987;  
Best Local Similarity 89.5%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagc 19  
||| ||||| |||||  
Db 747 TCGGCAGGTGGAGGTCAGC 765

RESULT 20  
US-08-184-009-194  
; Sequence 194, Application US/08184009  
; Patent No. 5833975  
; GENERAL INFORMATION:  
; APPLICANT: Paoletti, Enzo  
; APPLICANT: Tartaglia, James  
; APPLICANT: Cox, William I.  
; TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY  
; NUMBER OF SEQUENCES: 217  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis, Morris & Safford  
; STREET: 530 Fifth Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/184,009  
FILING DATE: 19-JAN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2530  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
TELEX: 425066CURTMS  
INFORMATION FOR SEQ ID NO: 194:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1018 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-184-009-194

Query Match 63.2%; Score 15.8; DB 2; Length 1018;  
Best Local Similarity 89.5%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagc 19  
||| ||||| |||||  
DB 778 TCGCAGGTGGAGGTGAGC 796

RESULT 21  
US-08-458-356-194  
Sequence 194, Application US/08458356  
Patent No. 5942235  
GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
APPLICANT: Tartaglia, James  
APPLICANT: Cox, William I.  
TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY  
NUMBER OF SEQUENCES: 217  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,356  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/184,009  
FILING DATE: 19-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2530  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
TELEX: 425066CURTMS

INFORMATION FOR SEQ ID NO: 194:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1018 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-458-356-194

Query Match 63.2%; Score 15.8; DB 2; Length 1018;  
Best Local Similarity 89.5%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagc 19  
||| ||||| |||||  
DB 778 TCGCAGGTGGAGGTGAGC 796

RESULT 22  
US-08-751-767A-3  
Sequence 3, Application US/08751767A  
Patent No. 5994104  
GENERAL INFORMATION:  
APPLICANT: ANDERSON, ROBERT J.  
APPLICANT: GRANT, HUGH  
APPLICANT: MACDONALD, IAN D.  
TITLE OF INVENTION: INTERLUKIN-12 FUSION PROTEIN  
NUMBER OF SEQUENCES: 80  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHVE P.C.  
STREET: 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/751,767A  
FILING DATE: 08-NOV-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: SADOFF, B.J.  
REGISTRATION NUMBER: 36,663  
REFERENCE/DOCKET NUMBER: 117-221  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 7038164091  
TELEFAX: 7038164100  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1399 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 43..1026  
US-08-751-767A-3

Query Match 63.2%; Score 15.8; DB 2; Length 1399;  
Best Local Similarity 89.5%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagc 19  
||| ||||| |||||  
DB 789 TCGCAGGTGGAGGTGAGC 807

RESULT 23  
US-08-751-767A-11  
; Sequence 11, Application US/08751767A  
; Patent No. 5994104  
; GENERAL INFORMATION:  
; APPLICANT: ANDERSON, ROBERT J.  
; APPLICANT: GRANT, HUGH  
; APPLICANT: MACDONALD, IAN D.  
; TITLE OF INVENTION: INTERLUKIN-12 FUSION PROTEIN  
; NUMBER OF SEQUENCES: 80  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHYE P.C.  
; STREET: 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22201  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/751.767A  
; FILING DATE: 08-NOV-1996  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SADOFF, B.J.  
; REGISTRATION NUMBER: 36,663  
; REFERENCE/DOCKET NUMBER: 117-221  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 7038164091  
; TELEFAX: 7038164100  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1560 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1557  
US-08-751-767A-11  
  
Query Match 63.2%; Score 15.8; DB 2; Length 1560;  
Best Local Similarity 89.5%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 tcgtcaggtgcaggtcagc 19  
||| ||||| ||||| |||||  
Db 684 TCGGCGAGGTGGAGTGCAGC 702  
  
RESULT 24  
US-08-751-767A-9  
; Sequence 9, Application US/08751767A  
; Patent No. 5994104  
; GENERAL INFORMATION:  
; APPLICANT: ANDERSON, ROBERT J.  
; APPLICANT: GRANT, HUGH  
; APPLICANT: MACDONALD, IAN D.  
; TITLE OF INVENTION: INTERLUKIN-12 FUSION PROTEIN  
; NUMBER OF SEQUENCES: 80  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHYE P.C.  
; STREET: 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22201

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/751.767A  
; FILING DATE: 08-NOV-1996  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SADOFF, B.J.  
; REGISTRATION NUMBER: 36,663  
; REFERENCE/DOCKET NUMBER: 117-221  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 7038164091  
; TELEFAX: 7038164100  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1623 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1620  
US-08-751-767A-9  
  
Query Match 63.2%; Score 15.8; DB 2; Length 1623;  
Best Local Similarity 89.5%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 tcgtcaggtgcaggtcagc 19  
||| ||||| ||||| |||||  
Db 747 TCGGCGAGGTGGAGTGCAGC 765  
  
RESULT 25  
US-08-265-087-1  
; Sequence 1, Application US/08265087  
; Patent No. 5571515  
; GENERAL INFORMATION:  
; APPLICANT: Scott, Phillip  
; APPLICANT: Trinchieri, Giorgio  
; TITLE OF INVENTION: Compositions and Methods for Use of  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howson and Howson  
; STREET: Spring House Corporate Center, PO Box 457  
; CITY: Spring House  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19477  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/265.087  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/229,282  
; FILING DATE: 18-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bak, Mary E.  
; REGISTRATION NUMBER: 31,215  
; REFERENCE/DOCKET NUMBER: WST51AUSA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-540-9206

TELEFAX: 215-540-5818  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2362 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 33..1016  
US-08-265-087-1

Query Match 63.2%; Score 15.8; DB 1; Length 2362;  
Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagc 19  
||| ||||| ||||| |||||  
DB 779 TCGGCAGGTGGAGGTCAGC 797

## RESULT 26

US-08-621-493-1  
; Sequence 1, Application US/08621493  
; Patent No. 5723127  
; GENERAL INFORMATION:  
; APPLICANT: Scott, Phillip  
; APPLICANT: Trinchieri, Giorgio  
; TITLE OF INVENTION: Compositions and Methods for Use of  
; TITLE OF INVENTION: IL-12 as an Adjuvant  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howson and Howson  
; STREET: Spring House Corporate Center, PO Box 457  
; CITY: Spring House  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19477

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/621,493  
; FILING DATE: 25-MAR-1996  
; CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/265,087  
; FILING DATE: 17-JUN-1994  
; APPLICATION NUMBER: US 08/229,282  
; FILING DATE: 18-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bak, Mary E.  
; REGISTRATION NUMBER: 31,215  
; REFERENCE/DOCKET NUMBER: WST51AUSA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-540-9206  
; TELEFAX: 215-540-5818

INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2362 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 33..1016

US-08-621-493-1

Query Match 63.2%; Score 15.8; DB 1; Length 2362;  
Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagc 19  
||| ||||| ||||| |||||  
DB 779 TCGGCAGGTGGAGGTCAGC 797

## RESULT 27

US-08-965-688-1  
; Sequence 1, Application US/08965688  
; Patent No. 5976539  
; GENERAL INFORMATION:  
; APPLICANT: Scott, Phillip  
; APPLICANT: Trinchieri, Giorgio  
; TITLE OF INVENTION: Compositions and Methods for Use of  
; TITLE OF INVENTION: IL-12 as an Adjuvant  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howson and Howson  
; STREET: Spring House Corporate Center, PO Box 457  
; CITY: Spring House  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19477

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/965,688  
; FILING DATE:

CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/621,493  
; FILING DATE: 25-MAR-1996  
; APPLICATION NUMBER: 08/265,087  
; FILING DATE: 17-JUN-1994  
; APPLICATION NUMBER: US 08/229,282  
; FILING DATE: 18-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215  
; REFERENCE/DOCKET NUMBER: WST51AUSA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-540-9206  
; TELEFAX: 215-540-5818  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2362 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 33..1016  
US-08-965-688-1

Query Match 63.2%; Score 15.8; DB 2; Length 2362;  
Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagc 19  
||| ||||| ||||| |||||  
DB 779 TCGGCAGGTGGAGGTCAGC 797

## RESULT 28



```

US-09-260-173-1
; Sequence 1, Application US/09260173
; Patent No. 6168923
; GENERAL INFORMATION:
; APPLICANT: Scott, Phillip
; Trinchieri, Giorgio
; TITLE OF INVENTION: Compositions and Methods for Use of
; IL-12 as an Adjuvant
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, PO Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/260,173
; FILING DATE: 01-Mar-1999
; CLASSIFICATION: <Unknown>
; 17-JUN-1994
; 18-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/965,688
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/265,087
; FILING DATE: 17-JUN-1994
; APPLICATION NUMBER: US 08/229,282
; FILING DATE: 18-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST51AUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9206
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2362 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 33..1016
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-260-173-1

```

```

Query Match      63.2%; Score 15.8; DB 4; Length 2362;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagc 19
   ||| ||||| |||||
DB 779 TCGGCAGGTGGAGGTGAGC 797

```

```

RESULT 29
US-08-471-033-17/c
; Sequence 17, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J

```

```

; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,033
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: P-40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SQLv3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2655 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..2652
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for 100 kd VIPIA(a) protein from AB78"
US-08-471-033-17

Query Match      63.2%; Score 15.8; DB 1; Length 2655;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 ggtgcaggtcagcagcttg 25
   ||||| ||||| |||||
DB 45 GGTGCAGGTGACCACGCTG 27

RESULT 30
US-08-471-033-26/c
; Sequence 26, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A

```

APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,033  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: P-40,403  
REFERENCE/DOCKET NUMBER: CSC 1695/CIP3/DIV7 - SOLV3  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2655 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Synthetic DNA"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..2655  
OTHER INFORMATION: /note= "maize optimized DNA"  
OTHER INFORMATION: sequence encoding VIPLA(a)"  
US-08-471-033-26

Query Match 63.2%; Score 15.8; DB 1; Length 2655;  
Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 7 ggtgcaggtcagcagcttg 25  
|||||  
Db 45 GGTGCAGGTCCACCGCTG 27

RESULT 31  
US-08-471-044-17/c  
Sequence 17, Application US/08471044  
Patent No. 5840868  
GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Koziel, Michael G

APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,044  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/463,483  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2655 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..2652  
OTHER INFORMATION: /note= "Maize optimized DNA"  
OTHER INFORMATION: sequence for 100 kd VIPLA(a) protein from AB78"  
US-08-471-044-17

Query Match 63.2%; Score 15.8; DB 2; Length 2655;  
Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 7 ggtgcaggtcagcagcttg 25  
|||||  
Db 45 GGTGCAGGTCCACCGCTG 27

RESULT 32  
US-08-471-044-26/c  
Sequence 26, Application US/08471044

Patent No. 5840868  
GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Koziel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,044  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/463,483  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SQLV3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2655 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Synthetic DNA"  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1..2655  
OTHER INFORMATION: /note= "maize optimized DNA"  
OTHER INFORMATION: sequence encoding VIPLA(a)"  
US-08-471-044-26

Query Match 63.2%; Score 15.8; DB 2; Length 2655;  
Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 ggtgcaggtcagcagcttg 25  
|||||  
DB 45 GGTGCAGGTCCACCGCTG 27

RESULT 33  
US-08-463-483A-17/c  
Sequence 17, Application US/08463483A  
Patent No. 5849870  
GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Koziel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: No. 5849870el Pesticidal Proteins and Strains  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/463,483A  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Spruill, W. Murray  
REGISTRATION NUMBER: 32,943  
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8615  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2655 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1..2652  
OTHER INFORMATION: /note= "Maize optimized DNA"  
OTHER INFORMATION: sequence for 100 kd VIPLA(a) protein from AB78"  
US-08-463-483A-17

Query Match 63.2%; Score 15.8; DB 2; Length 2655;  
Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 ggtgcaggtcagcagcttg 25  
|||||  
DB 45 GGTGCAGGTCCACCGCTG 27

## RESULT 34

US-08-463-483A-26/c  
; Sequence 26, Application US/08463483A  
; Patent No. 5849870  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Gregory W  
; APPLICANT: Koziel, Michael G  
; APPLICANT: Mullins, Martha A  
; APPLICANT: Nye, Gordon J  
; APPLICANT: Carr, Brian  
; APPLICANT: Desai, Nalini M  
; APPLICANT: Kostichka, N. Kristy  
; APPLICANT: Duck, Nicholas B  
; APPLICANT: Estruch, Juan J  
; TITLE OF INVENTION: No. 5849870el Pesticidal Proteins and Strains  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30B  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/463.483A  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/314,594  
; FILING DATE: 09-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/218,018  
; FILING DATE: 23-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/037,057  
; FILING DATE: 25-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Spruill, W. Murray  
; REGISTRATION NUMBER: 32,943  
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8615  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2655 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "Synthetic DNA"  
; HYPOTHETICAL: NO  
; FEATURE:  
; NAME/KEY: misc-feature  
; LOCATION: 1..2655  
; OTHER INFORMATION: /note= "maize optimized DNA"  
; OTHER INFORMATION: sequence encoding VIP1A(a)"  
US-08-463-483A-26

Query Match 63.2%; Score 15.8; DB 2; Length 2655;

Best Local Similarity 89.5%; Pred. No. 1.5e-02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 ggtgcaggtcagcacgttg 25

|||||||

Db 45 GGTGCAGGTCAACGCTG 27

## RESULT 35

US-08-471-046A-17/c  
; Sequence 17, Application US/08471046A  
; Patent No. 5866326  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Gregory W  
; APPLICANT: Koziel, Michael G  
; APPLICANT: Mullins, Martha A  
; APPLICANT: Nye, Gordon J  
; APPLICANT: Carr, Brian  
; APPLICANT: Desai, Nalini M  
; APPLICANT: Kostichka, N. Kristy  
; APPLICANT: Duck, Nicholas B  
; APPLICANT: Estruch, Juan J  
; TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5866326artis Corporation  
; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park  
; STATE: NC  
; COUNTRY: USA  
; ZIP: 27709  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30B  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/471,046A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/463,483  
; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/314,594  
; FILING DATE: 09-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/218,018  
; FILING DATE: 23-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/037,057  
; FILING DATE: 25-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241  
; REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SQLv4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8587  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2655 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: misc-feature  
; LOCATION: 1..2652  
; OTHER INFORMATION: /note= "Maize optimized DNA"  
; OTHER INFORMATION: sequence for 100 kd VIP1A(a) protein from AB78"  
US-08-471-046A-17

Query Match

63.2%; Score 15.8; DB 2; Length 2655;

Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 7 ggtgcagggtcagcagcttg 25  
|||||  
Db 45 GGTGCAGGTCCACCACGCTG 27

## RESULT 36

US-08-471-046A-26/c  
; Sequence 26, Application US/08471046A  
; Patent No. 5866326

## GENERAL INFORMATION:

; APPLICANT: Warren, Gregory W  
; APPLICANT: Koziel, Michael G  
; APPLICANT: Mullins, Martha A  
; APPLICANT: Nye, Gordon J  
; APPLICANT: Carr, Brian  
; APPLICANT: Desai, Nalini M  
; APPLICANT: Kostichka, N. Kristy  
; APPLICANT: Duck, Nicholas B  
; APPLICANT: Estruch, Juan J  
; TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal  
; TITLE OF INVENTION: Protein Genes  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5866326artis Corporation  
; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park  
; STATE: NC  
; COUNTRY: USA  
; ZIP: 27709

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B.  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/471,046A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION NUMBER: US 08/463,483  
; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/314,594  
; FILING DATE: 09-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/218,018  
; FILING DATE: 23-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/037,057  
; FILING DATE: 25-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241  
; REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLV4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8587  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2655 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "Synthetic DNA"  
; HYPOTHETICAL: NO  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..2655  
; OTHER INFORMATION: /note= "maize optimized DNA"

; OTHER INFORMATION: sequence encoding VIPLA(a)"  
; US-08-471-046A-26

Query Match 63.2%; Score 15.8; DB 2; Length 2655;  
Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 7 ggtgcagggtcagcagcttg 25  
|||||  
Db 45 GGTGCAGGTCCACCACGCTG 27

## RESULT 37

US-08-470-566B-17/c  
; Sequence 17, Application US/08470566B  
; Patent No. 5872212

## GENERAL INFORMATION:

; APPLICANT: Warren, Gregory W  
; APPLICANT: Koziel, Michael G  
; APPLICANT: Mullins, Martha A  
; APPLICANT: Nye, Gordon J  
; APPLICANT: Carr, Brian  
; APPLICANT: Desai, Nalini M  
; APPLICANT: Kostichka, N. Kristy  
; APPLICANT: Duck, Nicholas B  
; APPLICANT: Estruch, Juan J  
; TITLE OF INVENTION: No. 5872212el Pesticidal Proteins and Strains  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5872212artis Corporation  
; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park  
; STATE: NC  
; COUNTRY: USA  
; ZIP: 27709

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/470,566B  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/463,483  
; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/314,594  
; FILING DATE: 09-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/218,018  
; FILING DATE: 23-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/037,057  
; FILING DATE: 25-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241  
; REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8587  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2655 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO

```
;
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..2652
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for 100 kd VIPIA(a) protein from AB78"
US-08-470-566B-17

Query Match      63.2%; Score 15.8; DB 2; Length 2655;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 ggtgcaggtcagcagcttg 25
   ||||| ||||| ||||| ||
Db 45 GGTGCAGGTCACCACGCTG 27

RESULT 38
US-08-470-566B-26/c
; Sequence 26, Application US/08470566B
; Patent No. 5872212
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5872212el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5872212artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,566B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CCG1695/CIP3/DIV4 - SOLV4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2655 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

```
;
;
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..2655
; OTHER INFORMATION: /note= "maize optimized DNA
; OTHER INFORMATION: sequence encoding VIPIA(a)"
US-08-470-566B-26

Query Match      63.2%; Score 15.8; DB 2; Length 2655;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 ggtgcaggtcagcagcttg 25
   ||||| ||||| ||||| ||
Db 45 GGTGCAGGTCACCACGCTG 27

RESULT 39
US-08-469-334-17/c
; Sequence 17, Application US/08469334
; Patent No. 5990383
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5990383el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,334
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/463,483
; FILING DATE:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8615
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
```

LENGTH: 2655 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..2652  
OTHER INFORMATION: /note= "Maize optimized DNA  
OTHER INFORMATION: sequence for 100 kd VIPIA(a) protein from AB78"  
US-08-469-334-17

Query Match 63.2%; Score 15.8; DB 2; Length 2655;  
Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 ggtgcaggtcagcacgttg 25  
||||||| |||||  
DB 45 GGTGAGGTCCACCGCTG 27

## RESULT 40

US-08-469-334-26/c  
; Sequence 26, Application US/08469334  
; Patent No. 5990383

; GENERAL INFORMATION:  
; APPLICANT: Warren, Gregory W

; APPLICANT: Koziel, Michael G  
; APPLICANT: Mullins, Martha A

; APPLICANT: Nye, Gordon J  
; APPLICANT: Carr, Brian

; APPLICANT: Desai, Nalini M  
; APPLICANT: Kostichka, N. Kristy

; APPLICANT: Duck, Nicholas B  
; APPLICANT: Estruch, Juan J

; TITLE OF INVENTION: No. 5990383el Pesticidal Proteins and Strains  
; NUMBER OF SEQUENCES: 50

; CORRESPONDENCE ADDRESS:  
; ADDRESS: CIBA-GEIGY Corporation

; STREET: 7 Skyline Drive  
; CITY: Hawthorne

; STATE: NY  
; COUNTRY: USA

; ZIP: 10532

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30B  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/469,334  
; FILING DATE: 06-JUN-1995

; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/463,483  
; FILING DATE:

; APPLICATION NUMBER: US 08/314,594  
; FILING DATE: 09-SEP-1994

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/218,018

; FILING DATE: 23-MAR-1994  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/037,057  
; FILING DATE: 25-MAR-1993

; ATTORNEY/AGENT INFORMATION:  
; NAME: SFRULL, W. MURRAY

; REGISTRATION NUMBER: 32,943  
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8615

TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2655 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "Synthetic DNA"  
; HYPOTHETICAL: NO  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..2655  
; OTHER INFORMATION: /note= "maize optimized DNA"  
; OTHER INFORMATION: sequence encoding VIPIA(a)"  
US-08-469-334-26

Query Match 63.2%; Score 15.8; DB 2; Length 2655;  
Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 ggtgcaggtcagcacgttg 25  
||||||| |||||  
DB 45 GGTGAGGTCCACCGCTG 27

## RESULT 41

US-09-300-529-17/c

; Sequence 17, Application US/09300529  
; Patent No. 6066783

; GENERAL INFORMATION:  
; APPLICANT: Warren, Gregory W

; APPLICANT: Koziel, Michael G  
; APPLICANT: Mullins, Martha A

; APPLICANT: Nye, Gordon J  
; APPLICANT: Carr, Brian

; APPLICANT: Desai, Nalini M  
; APPLICANT: Kostichka, N. Kristy

; APPLICANT: Duck, Nicholas B  
; APPLICANT: Estruch, Juan J

; TITLE OF INVENTION: Genes Encoding Insecticidal Proteins  
; NUMBER OF SEQUENCES: 50

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6066783artis Corporation

; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park

; STATE: NC  
; COUNTRY: USA

; ZIP: 27709

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30B  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/300,529  
; FILING DATE: TBA

; CLASSIFICATION:  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/469,334  
; FILING DATE: 06-JUN-1995

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/463,483

; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/314,594  
; FILING DATE: 09-SEP-1994

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/218,018

; FILING DATE: 23-MAR-1994  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/037,057

```

; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: S-19506L
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2655 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..2652
; OTHER INFORMATION: /note= "Maize optimized DNA
; sequence for 100 kd VIPIA(a) protein from AB78"
US-09-300-529-17

Query Match 63.2%; Score 15.8; DB 3; Length 2655;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 ggtgcaggtcagcacgttg 25
Db 45 GGTGCAGGTCACCCAGCTG 27

RESULT 42
US-09-300-529-26/c
; Sequence 26, Application US/09300529
; Patent No. 6066783
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: Genes Encoding Insecticidal Proteins
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6066783artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.308
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/300,529
; FILING DATE: TBA
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/469,334
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA: US 08/218,018
; APPLICATION NUMBER: 38,241
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: S-19506L
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2655 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..2655
; OTHER INFORMATION: /note= "maize optimized DNA
; sequence encoding VIPIA(a)"
US-09-300-529-26

Query Match 63.2%; Score 15.8; DB 3; Length 2655;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 ggtgcaggtcagcacgttg 25
Db 45 GGTGCAGGTCACCCAGCTG 27

RESULT 43
US-08-751-767A-7
; Sequence 7, Application US/08751767A
; Patent No. 5994104
; GENERAL INFORMATION:
; APPLICANT: ANDERSON, ROBERT J.
; APPLICANT: GRANT, HUGH
; APPLICANT: MACDONALD, IAN D.
; TITLE OF INVENTION: INTERLUKIN-12 FUSION PROTEIN
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHUYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,767A
; FILING DATE: 08-NOV-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 117-221
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 7038164091

```



TELEFAX: 7038164100  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6139 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 3222..4841  
US-08-751-767A-7

Query Match 63.2%; Score 15.8; DB 2; Length 6139;  
Best Local Similarity 89.5%; Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tcgtcaggtgcaggtcagc 19  
||| ||||| ||||| |||||  
Db 3968 TCGGCAGGTGGAGTCAGC 3986

## RESULT 44

US-08-927-219-36/c  
; Sequence 36, Application US/08927219  
; Patent No. 6187533

; GENERAL INFORMATION:  
; APPLICANT: Bell, Graeme I.

; APPLICANT: Yamagata, Kazuya

; APPLICANT: Oda, Naohisha

; APPLICANT: Kalsaki, Pamela J.

; APPLICANT: Furuta, Hiroto

; APPLICANT: Horikawa, Yukio

; APPLICANT: Menzel, Stephen

; TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY

; TITLE OF INVENTION: GENES HEPATOCYTE NUCLEAR FACTOR (HNF) 1 ALPHA, HNF-1BETA

; NUMBER OF SEQUENCES: 147

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P.O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: USA

; ZIP: 77210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/927,219

; FILING DATE: Concurrently Herewith

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/029,679

; FILING DATE: 30-OCT-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/028,056

; FILING DATE: 02-OCT-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/025,719

; FILING DATE: 10-SEP-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Wilson, Mark B.

; REGISTRATION NUMBER: 37,259

; REFERENCE/DOCKET NUMBER: ARCD:272

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 512/418-3000

; TELEFAX: 512/474-7577

; INFORMATION FOR SEQ ID NO: 36:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 796 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: join(286...312, 316...375)  
US-08-927-219-36

Query Match 62.4%; Score 15.6; DB 4; Length 796;  
Best Local Similarity 81.8%; Pred. No. 1.7e+02;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 cgtcaggtgcaggtcagcagt 23  
||| ||||| ||||| |||||  
Db 640 CGGCAGGGCGAGTCAGCACCT 619

## RESULT 45

US-08-767-942A-12/c

; Sequence 12, Application US/08767942A

; Patent No. 6068982

; GENERAL INFORMATION:

; APPLICANT: Rolfe, Mark

; APPLICANT: Chiu, M. Isabel

; APPLICANT: Berlin, Vivian

; APPLICANT: Damagnez, Veronique

; APPLICANT: Draetta, Giulio

; APPLICANT: Guillaume, Cottarel

; TITLE OF INVENTION: UBIQUITIN CONJUGATING ENZYMES

; NUMBER OF SEQUENCES: 45

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FOLEY, HOAG & ELIOT LLP

; STREET: One Post Office Square

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02109-2170

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/767,942A

; FILING DATE: 17-DEC-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Vincent, Matthew P.

; REGISTRATION NUMBER: 36,709

; REFERENCE/DOCKET NUMBER: MIV-029.04

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-832-1000

; TELEFAX: 617-832-7000

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 907 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 34..507

US-08-767-942A-12

Query Match 62.4%; Score 15.6; DB 3; Length 907;  
Best Local Similarity 78.3%; Pred. No. 1.7e+02;  
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 gtcaggtgcaggtcagcagcttg 25  
||| ||||| ||||| |||||

Db 883 GNCAAGTGCAGCTCTGCACGGTG 861

Search completed: October 9, 2001, 11:40:08  
Job time: 1923 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: October 9, 2001, 13:46:52 ; Search time 5930.9 seconds  
(without alignments)  
39.846 Million cell updates/sec

Title: US-09-396-196F-10  
Perfect score: 25  
Sequence: 1 tcgtcaggtgcaggtcacgcttg 25

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues  
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

- 1: gb\_estl1:\*
- 2: gb\_est2:\*
- 3: gb\_est3:\*
- 4: gb\_est4:\*
- 5: gb\_est5:\*
- 6: gb\_est6:\*
- 7: gb\_est7:\*
- 8: gb\_est8:\*
- 9: gb\_est9:\*
- 10: gb\_est10:\*
- 11: gb\_est11:\*
- 12: gb\_est12:\*
- 13: gb\_est13:\*
- 14: gb\_est14:\*
- 15: gb\_est15:\*
- 16: gb\_est16:\*
- 17: gb\_est17:\*
- 18: gb\_est18:\*
- 19: gb\_est19:\*
- 20: gb\_est20:\*
- 21: gb\_est21:\*
- 22: gb\_est22:\*
- 23: gb\_est23:\*
- 24: gb\_est24:\*
- 25: gb\_est33:\*
- 26: gb\_est34:\*
- 27: gb\_est35:\*
- 28: gb\_est36:\*
- 29: gb\_est37:\*
- 30: gb\_est38:\*
- 31: gb\_est39:\*
- 32: gb\_est40:\*
- 33: em\_estba:\*
- 34: em\_estfun:\*
- 35: em\_esthum1:\*
- 36: em\_esthum2:\*
- 37: em\_esthum3:\*
- 38: em\_esthum4:\*
- 39: em\_esthum5:\*
- 40: em\_esthum6:\*
- 41: em\_esthum7:\*
- 42: em\_esthum8:\*
- 43: em\_esthum9:\*

44: em\_esthum10:\*

45: em\_esthum11:\*

46: em\_esthum12:\*

47: em\_esthum13:\*

48: em\_esthum14:\*

49: em\_esthum15:\*

50: em\_esthum16:\*

51: em\_esthum17:\*

52: em\_esthum18:\*

53: em\_esthum19:\*

54: em\_esthum20:\*

55: em\_esthum21:\*

56: em\_esthum22:\*

57: em\_esthum23:\*

58: em\_esthum24:\*

59: em\_esthum25:\*

60: em\_esthum26:\*

61: em\_esthum27:\*

62: em\_esthum28:\*

63: em\_estin1:\*

64: em\_estin2:\*

65: em\_estin3:\*

66: em\_estin4:\*

67: em\_estin5:\*

68: em\_estom1:\*

69: em\_estom2:\*

70: em\_estov1:\*

71: em\_estov2:\*

72: em\_estp11:\*

73: em\_estp12:\*

74: em\_estp13:\*

75: em\_estp14:\*

76: em\_estp15:\*

77: em\_estp16:\*

78: em\_estp17:\*

79: em\_estp18:\*

80: em\_estp19:\*

81: em\_estp110:\*

82: em\_estrol:\*

83: em\_estrol2:\*

84: em\_estrol3:\*

85: em\_estrol4:\*

86: em\_estrol5:\*

87: em\_estrol6:\*

88: em\_estrol7:\*

89: em\_estrol8:\*

90: em\_estrol9:\*

91: em\_estrol10:\*

92: em\_estrol11:\*

93: em\_estrol12:\*

94: em\_estrol13:\*

95: em\_estrol14:\*

96: em\_estrol15:\*

97: em\_estrol16:\*

98: em\_estrol17:\*

99: em\_estrol18:\*

100: em\_estrol19:\*

101: em\_estrol20:\*

102: gb\_est25:\*

103: gb\_est26:\*

104: gb\_est27:\*

105: gb\_est28:\*

106: gb\_est29:\*

107: gb\_est30:\*

108: gb\_est31:\*

109: gb\_est32:\*

110: gb\_est41:\*

111: gb\_est42:\*

112: gb\_est43:\*

113: gb\_est44:\*

114: gb\_est45:\*

115: gb\_est46:\*

116: gb\_est47:\*

117: gb\_est48: \*  
118: gb\_est49: \*  
119: gb\_est50: \*  
120: gb\_est51: \*  
121: gb\_est52: \*  
122: gb\_est53: \*  
123: gb\_est54: \*  
124: gb\_est55: \*  
125: gb\_est56: \*  
126: gb\_est57: \*  
127: gb\_est58: \*  
128: gb\_est59: \*  
129: gb\_est60: \*  
130: gb\_est61: \*  
131: gb\_est62: \*  
132: gb\_est63: \*  
133: gb\_est64: \*  
134: gb\_est65: \*  
135: gb\_est66: \*  
136: gb\_est67: \*  
137: gb\_est68: \*  
138: gb\_est69: \*  
139: gb\_est70: \*  
140: gb\_est71: \*  
141: gb\_est72: \*  
142: gb\_est73: \*  
143: gb\_est74: \*  
144: gb\_est75: \*  
145: gb\_est76: \*  
146: gb\_est77: \*  
147: gb\_est78: \*  
148: gb\_est79: \*  
149: gb\_est80: \*  
150: gb\_est81: \*  
151: gb\_est82: \*  
152: gb\_est83: \*  
153: gb\_est84: \*  
154: gb\_est85: \*  
155: gb\_est86: \*  
156: gb\_est87: \*  
157: gb\_est88: \*  
158: gb\_est89: \*  
159: gb\_est90: \*  
160: gb\_est91: \*  
161: gb\_est92: \*  
162: gb\_est93: \*  
163: gb\_est94: \*  
164: gb\_est95: \*  
165: gb\_est96: \*  
166: gb\_est97: \*  
167: gb\_est98: \*  
168: gb\_est99: \*  
169: gb\_est100: \*  
170: gb\_est101: \*  
171: gb\_est102: \*  
172: gb\_est103: \*  
173: gb\_est104: \*  
174: gb\_est105: \*  
175: gb\_est106: \*  
176: gb\_est107: \*  
177: gb\_est108: \*  
178: gb\_est109: \*  
179: gb\_est110: \*  
180: gb\_est111: \*  
181: gb\_est112: \*  
182: gb\_est113: \*  
183: gb\_est114: \*  
184: gb\_est115: \*  
185: gb\_est116: \*  
186: gb\_est117: \*  
187: gb\_est118: \*  
188: gb\_est119: \*  
189: gb\_est120: \*

190: gb\_est121: \*  
191: gb\_est122: \*  
192: gb\_est123: \*  
193: gb\_est124: \*  
194: gb\_est125: \*  
195: gb\_est126: \*  
196: gb\_est127: \*  
197: gb\_est128: \*  
198: gb\_est129: \*  
199: gb\_est130: \*  
200: gb\_est131: \*  
201: gb\_est132: \*  
202: gb\_est133: \*  
203: gb\_est134: \*  
204: gb\_est135: \*  
205: gb\_est136: \*  
206: gb\_est137: \*  
207: gb\_est138: \*  
208: gb\_est139: \*  
209: gb\_est140: \*  
210: gb\_est141: \*  
211: gb\_est142: \*  
212: gb\_est143: \*  
213: gb\_est144: \*  
214: gb\_est145: \*  
215: gb\_est146: \*  
216: gb\_est147: \*  
217: gb\_est148: \*  
218: gb\_est149: \*  
219: gb\_est150: \*  
220: gb\_est151: \*  
221: gb\_est152: \*  
222: gb\_est153: \*  
223: gb\_est154: \*  
224: gb\_est155: \*  
225: gb\_est156: \*  
226: gb\_est157: \*  
227: gb\_est158: \*  
228: gb\_est159: \*  
229: gb\_est160: \*  
230: gb\_est161: \*  
231: gb\_est162: \*  
232: gb\_est163: \*  
233: gb\_est164: \*  
234: gb\_est165: \*  
235: gb\_est166: \*  
236: gb\_est167: \*  
237: gb\_est168: \*  
238: gb\_est169: \*  
239: gb\_est170: \*  
240: gb\_est171: \*  
241: gb\_est172: \*  
242: gb\_est173: \*  
243: gb\_est174: \*  
244: gb\_est175: \*  
245: gb\_est176: \*  
246: gb\_est177: \*  
247: gb\_est178: \*  
248: gb\_est179: \*  
249: gb\_est180: \*  
250: gb\_est181: \*  
251: gb\_est182: \*  
252: gb\_est183: \*  
253: gb\_est184: \*  
254: gb\_est185: \*  
255: gb\_est186: \*  
256: gb\_est187: \*  
257: gb\_est188: \*  
258: gb\_est189: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.





```

/clone="AT28277"
/clone_lib="AT Drosophila melanogaster adult testes pOTB7"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="Plates AT.10-AT.120: DH5-alpha. Plates
AT.121-AT.319: DH5-alpha Tona"
/notes="Organ: ADULT testes; Vector: pOTB7; Site.1: EcoRI;
Site.2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into pOTB7. Plasmid cDNA library."
BASE COUNT      163 a      183 g      150 t
ORIGIN

```

```

Query Match      74.4%; Score 18.6; DB 149; Length 688;
Best Local Similarity 84.0%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagcagcttg 25
    ||| ||||| ||||| ||| |||
Db 255 TCGACAGGTGCAGGTCTGCAAGGTG 279

```

```

RESULT 5
BF504829
LOCUS
DEFINITION
BF504829.1 GI:11588130
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster
fruit fly.
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 697)
Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E., Berman
, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan
, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P.,
Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J.,
Park, S., Paragas, V., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E.,
Celisner, S. and Rubin, G.M.
Berkeley Drosophila Gene Collection Project
Unpublished (2000)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic sequence AB003582
Plate: AT 68 row: E column: 12
High quality sequence stop: 568.
Location/Qualifiers
1..697
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="AT06860"
/clone_lib="AT Drosophila melanogaster adult testes pOTB7"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="Plates AT.10-AT.120: DH5-alpha. Plates
AT.121-AT.319: DH5-alpha Tona"
/notes="Organ: ADULT testes; Vector: pOTB7; Site.1: EcoRI;
Site.2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into pOTB7. Plasmid cDNA library."

```

```

BASE COUNT      163 a      183 g      150 t
ORIGIN

```

```

Query Match      74.4%; Score 18.6; DB 149; Length 688;
Best Local Similarity 84.0%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagcagcttg 25
    ||| ||||| ||||| ||| |||
Db 255 TCGACAGGTGCAGGTCTGCAAGGTG 279

```

```

RESULT 5
BF504829
LOCUS
DEFINITION
BF504829.1 GI:11588130
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster
fruit fly.
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 697)
Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E., Berman
, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan
, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P.,
Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J.,
Park, S., Paragas, V., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E.,
Celisner, S. and Rubin, G.M.
Berkeley Drosophila Gene Collection Project
Unpublished (2000)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic sequence AB003582
Plate: AT 68 row: E column: 12
High quality sequence stop: 568.
Location/Qualifiers
1..697
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="AT06860"
/clone_lib="AT Drosophila melanogaster adult testes pOTB7"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="Plates AT.10-AT.120: DH5-alpha. Plates
AT.121-AT.319: DH5-alpha Tona"
/notes="Organ: ADULT testes; Vector: pOTB7; Site.1: EcoRI;
Site.2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into pOTB7. Plasmid cDNA library."

```

```

BASE COUNT      159 a      200 c      184 g      153 t      1 others
ORIGIN

```

```

Query Match      74.4%; Score 18.6; DB 149; Length 697;
Best Local Similarity 84.0%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagcagcttg 25
    ||| ||||| ||||| ||| |||
Db 234 TCGACAGGTGCAGGTCTGCAAGGTG 258

```

```

RESULT 6
BG480906/c
LOCUS
DEFINITION
BG480906.1 GI:13413185
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 765)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM1439 row: j column: 13
High quality sequence stop: 621.
Location/Qualifiers
1..765
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4653420"
/clone_lib="NIH_MGC_21"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: placenta; Vector: pOTB7; Site.1: XhoI;
Site.2: EcoRI; cDNA made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

```

```

BASE COUNT      182 a      191 c      224 g      168 t
ORIGIN

```

```

Query Match      74.4%; Score 18.6; DB 154; Length 765;
Best Local Similarity 84.0%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagcagcttg 25
    ||||| ||||| ||||| |||
Db 163 TCGTCAGGTGAGAGTCAGCACCTTG 139

```

```

RESULT 7
BF313972/c
LOCUS
DEFINITION
BF313972.1 GI:13413185
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 765)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM1439 row: j column: 13
High quality sequence stop: 621.
Location/Qualifiers
1..765
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4653420"
/clone_lib="NIH_MGC_21"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: placenta; Vector: pOTB7; Site.1: XhoI;
Site.2: EcoRI; cDNA made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

```

```

BASE COUNT      182 a      191 c      224 g      168 t
ORIGIN

```

```

Query Match      74.4%; Score 18.6; DB 154; Length 765;
Best Local Similarity 84.0%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcgtcaggtgcaggtcagcagcttg 25
    ||||| ||||| ||||| |||
Db 163 TCGTCAGGTGAGAGTCAGCACCTTG 139

```

```

RESULT 7
BF313972/c
LOCUS
DEFINITION
BF313972.1 GI:13413185
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 765)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM1439 row: j column: 13
High quality sequence stop: 621.
Location/Qualifiers
1..765
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4653420"
/clone_lib="NIH_MGC_21"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: placenta; Vector: pOTB7; Site.1: XhoI;
Site.2: EcoRI; cDNA made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

```





DEFINITION e3e05ne.r1 Neurospora crassa evening cDNA library Neurospora crassa  
cDNA clone e3e05ne 3', mRNA sequence.

ACCESSION AW710402  
VERSION AW710402.1 GI:7599490  
KEYWORDS EST.  
SOURCE Neurospora crassa.

ORGANISM Neurospora crassa  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Sordariales; Sordariaceae; Neurospora.

REFERENCE 1 (bases 1 to 318)

AUTHORS Zhu, H., Lai, H., Kupfer, D., Dunlap, J.C. and Roe, B.A.

TITLE Two Neurospora crassa EST Databases

JOURNAL Unpublished (1998)

COMMENT Other\_ESTS: e3e05ne.fl

Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu

Department of Chemistry and Biochemistry

Advanced Center for Genome Technology, University of Oklahoma

620 Parrington Oval, Norman, OK 73019, USA

Tel: 405 325 4912

Fax: 405 325 7762

Email: broe@ou.edu

Seq primer: Universal Reverse Primer

High quality sequence stop: 80.

Location/Qualifiers

1. 318

/organism="Neurospora crassa"

/strain="Strain 30-7 (bd; A)"

/db\_xref="taxon:5141"

/clone="e3e05ne"

/tissue\_type="tissue harvested following 22hr growth in

dark"

/note="Vector: pBluescript SK-; Site\_1: XbaI; Site\_2:

EcoRI; See: Bell-Petersen, D., et al. PNAS 93:13096,1996.

5' end of cDNA cloned into XbaI site of pBluescript; 3'

end of cDNA cloned into EcoRI site of pBluescript"

BASE COUNT 62 a 81 c 67 g 108 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 119; Length 318;

Best Local Similarity 87.0%; Pred. NO. 4.7e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cgtcagggtcaggtcagcagctt 24

||||| ||||| ||||| ||||| |||||

Db 99 CGTCAGGTCGCTGCTAGCACGTT 121

RESULT 10

AA809928/c

LOCUS AA809928 461 bp mRNA EST 19-FEB-1998

DEFINITION oa93g06.s1 NCI-CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:1319866 3'

similar to gb:X74301\_cds1 MHC CLASS II TRANSACTIVATOR CIITA (HUMAN

); mRNA sequence.

ACCESSION AA809928

VERSION AA809928.1 GI:2879334

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 461)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,

Ph.D., Gerald Marti, M.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 904 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 221.

Location/Qualifiers

1. 461

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1319866"

/clone\_lib="NCI-CGAP\_GCB1"

/tissue\_type="germinal center B cell"

/lab\_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA

was prepared from human tonsillar cells enriched for

germinal center B cells by flow sorting (CD20+, IgD+),

provided by Dr. Louis M. Staudt (NCI), Dr. David Allman

(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was

primed with a Not I - oligo(GT) primer

[5'-TGTTACCAATCTGAAGTGGAGCGCGCCCTCATTTTTTTTTTTT-3'

]. Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 94 a 140 c 133 g 93 t 1 others

ORIGIN

QY 1 tcgtcagggtcaggtcagcagct 23

||||| ||||| ||||| ||||| |||||

Db 276 TCGTCCGGTCCGCTCCGACGT 254

RESULT 11

AW710401/c

LOCUS AW710401 480 bp mRNA EST 25-APR-2000

DEFINITION e3e05ne.fl Neurospora crassa evening cDNA library Neurospora crassa

cDNA clone e3e05ne 5', mRNA sequence.

ACCESSION AW710401

VERSION AW710401.1 GI:7599489

KEYWORDS EST.

SOURCE Neurospora crassa.

ORGANISM Neurospora crassa

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Sordariales; Sordariaceae; Neurospora.

REFERENCE 1 (bases 1 to 480)

AUTHORS Zhu, H., Lai, H., Kupfer, D., Dunlap, J.C. and Roe, B.A.

TITLE Two Neurospora crassa EST Databases

JOURNAL Unpublished (1998)

COMMENT Other\_ESTS: e3e05ne.fl

Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu

Department of Chemistry and Biochemistry

Advanced Center for Genome Technology, University of Oklahoma

620 Parrington Oval, Norman, OK 73019, USA

Tel: 405 325 4912

Fax: 405 325 7762

Email: broe@ou.edu

We anticipate the future release of the cDNA clones to the Fungal

Genetics Stock Center

Seq primer: Universal Forward Primer

High quality sequence stop: 401.

Location/Qualifiers

1. 480

/organism="Neurospora crassa"

```

/strain="Strain 30-7 (bd; A)"
/db_xref="taxon:5141"
/clone="e3e05ne"
/clone_lib="Neurospora crassa evening cDNA library"
/tissue_type="tissue harvested following 22hr growth in
dark"
/note="Vector: pBluescript SK-; Site_1: XbaI; Site_2:
EcoRI; See: Bell-Pedersen,D., et al. PNAS 93:13096,1996.
5' end of cDNA cloned into XbaI site of pBluescript; 3'
end of cDNA cloned into EcoRI site of pBluescript"
BASE COUNT      146 a 121 c 127 g 86 t
ORIGIN

```

```

Query Match      72.8%; Score 18.2; DB 119; Length 480;
Best Local Similarity 87.0%; Pred. No. 4.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  2  cgtcagggtcaggtcagcagctt 24
      |||||
Db   431  CGTCAGGTGCTGCTTAGCAGGTT 409

```

```

RESULT 12
A2656921/c
LOCUS      A2656921      726 bp      DNA      GSS      14-DEC-2000
DEFINITION IM0532H13R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0532H13 R, DNA sequence.
ACCESSION  A2656921
VERSION     A2656921.1 GI:11794067
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 726)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
TITLE       Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL     Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0532 row: H column: 13
            Seq primer: CACACAGGAACACGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 726.
            Location/Qualifiers
                1..726
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC1M0532H13"
                /clone_lib="Mouse 10kb plasmid UUGC1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                /note="Vector: PWD42nv; Purified genomic DNA from M.
            musculus C57BL/6J (male) was obtained from the Jackson
            Laboratory Mouse DNA Resource
            (http://www.jax.org/resources/documents/dnares/). The DNA
            was hydrodynamically sheared by repeated passage through a
            0.005 inch orifice at constant velocity. The sheared DNA
            was blunt end-repaired with T4 DNA polymerase and T4
            polynucleotide kinase. Adaptor oligonucleotides were

```

```

FEATURES
    source
        1..726
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /clone="UUGC1M0532H13"
        /clone_lib="Mouse 10kb plasmid UUGC1M library"
        /sex="Male"
        /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
        /note="Vector: PWD42nv; Purified genomic DNA from M.
    musculus C57BL/6J (male) was obtained from the Jackson
    Laboratory Mouse DNA Resource
    (http://www.jax.org/resources/documents/dnares/). The DNA
    was hydrodynamically sheared by repeated passage through a
    0.005 inch orifice at constant velocity. The sheared DNA
    was blunt end-repaired with T4 DNA polymerase and T4
    polynucleotide kinase. Adaptor oligonucleotides were

```

```

ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gil4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      207 a 187 c 143 g 189 t
ORIGIN

```

```

Query Match      72.8%; Score 18.2; DB 247; Length 726;
Best Local Similarity 87.0%; Pred. No. 4.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  3  gtcagggtcaggtcagcagcttg 25
      |||||
Db   674  GTCAGGTGAAGTCAAGACATTG 652

```

```

RESULT 13
CNS02BOY
LOCUS      CNS02BOY      1041 bp      DNA      GSS      12-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
253N02 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION  AL190123
VERSION     AL190123.1 GI:7828227
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
ORGANISM    Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1 (bases 1 to 1041)
AUTHORS     Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
TITLE       Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 1041)
AUTHORS     Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
TITLE       Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
JOURNAL     Unpublished
REFERENCE   3 (bases 1 to 1041)
AUTHORS     Direct Submission
TITLE       Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
            Location/Qualifiers
                1..1041
                /organism="Tetraodon nigroviridis"
                /db_xref="taxon:99883"
                /clone="253N02"
                /clone_lib="G"
                /note="Genoscope sequence ID : C0AG253DG01LPL1-end : T7"
BASE COUNT      253 a 301 c 285 g 180 t 22 others
ORIGIN

```

```

Query Match      72.0%; Score 18; DB 220; Length 1041;

```

Best Local Similarity 90.0%; Pred. No. 6.2e+02;  
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 cgtcaggtgcaggtcagcac 21  
| | | | | | | | | | | | | | | | | | | | | |  
Db 92 CVCAGGTGAGGTAGCAC 111

RESULT 14  
A2790140/c  
LOCUS A2790140 528 bp DNA GSS 16-FEB-2001  
DEFINITION 2M0038E01R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0038E01 R, DNA sequence.  
ACCESSION A2790140  
VERSION A2790140.1 GI:12931689  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 528)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.  
REFERENCE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0038 row: E column: 01  
Seq primer: CACAGGAACACGTATGACC  
Class: plasmid ends  
High quality sequence stop: 528.

FEATURES  
source  
1..528  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0038E01"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi14732114|gbIA129072.1), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 154 a 111 c 114 g 149 t  
ORIGIN

Query Match 71.2%; Score 17.8; DB 249; Length 528;  
Best Local Similarity 90.5%; Pred. No. 7.3e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 tcaggtgcaggtcagcacgtt 24  
| | | | | | | | | | | | | | | | | | | | | |  
Db 366 TCAGGTGAGGTGAGCACGAT 346

RESULT 15  
A1478077/c  
LOCUS A1478077 730 bp mRNA EST 09-AUG-1999  
DEFINITION MPMGP621\_042D15 MPMGP621 Strongylocentrotus purpuratus cDNA clone 042D15 similar to beta-catenin, mRNA sequence.  
ACCESSION A1478077  
VERSION A1478077.1 GI:5713397  
KEYWORDS EST.  
SOURCE Strongylocentrotus purpuratus.  
ORGANISM Strongylocentrotus purpuratus  
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoida;  
Strongylocentrotidae; Strongylocentrotus.  
1 (bases 1 to 730)  
Poustka, A.J., Herwig, R., Krause, A., Hennig, S., Meier-Ewert, S. and Lehrach, H.  
REFERENCE Towards the gene catalogue of sea urchin development: The construction and analysis of an unfertilized egg cDNA library highly normalized by oligonucleotide fingerprinting  
Genomics 59 (2), 122-133 (1999)  
99339976  
Contact: Poustka A.J.  
Department Lehrach  
Max-Planck-Institute for Molecular Genetics  
Innestrasse 73, 14195 Berlin, Germany  
Tel: +49-30-84131623  
Fax: +49-30-84131128  
Email: poustka@mpg-berlin-dahlem.mpg.de  
Insert Length: 1500 Std Error: 0.00  
High quality sequence stop: 53.  
High quality sequence stop: 541.

FEATURES  
source  
1..730  
Location/Qualifiers  
/organism="Strongylocentrotus purpuratus"  
/db\_xref="taxon:7668"  
/clone="042D15"  
/clone\_lib="MPMGP621"  
/tissue\_type="unfertilized egg"  
/note="Vector: pSport1; oligo(dT) primed, directionally cloned library"

BASE COUNT 167 a 208 c 165 g 182 t 8 others  
ORIGIN

Query Match 71.2%; Score 17.8; DB 20; Length 730;  
Best Local Similarity 90.5%; Pred. No. 7.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 caggtgcaggtcagcacgttg 25  
| | | | | | | | | | | | | | | | | | | | | |  
Db 370 CAGGCGAGGTGAGCACGTTG 350

RESULT 16  
BF081725/c  
LOCUS BF081725 407 bp mRNA EST 18-OCT-2000  
DEFINITION PM0-AN0087-130900-003-all AN0087 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF081725  
VERSION BF081725.1 GI:10875555  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE 1 (bases 1 to 407)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
          Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
          Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
          Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
          ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
          Simpson,A.J.
TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
          sequence tags
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE   20202663
COMMENT   Contact: Simpson A.J.G.
          Laboratory of Cancer Genetics
          Ludwig Institute for Cancer Research
          Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
          Brazil
          Tel: +55-11-2704922
          Fax: +55-11-2707001
          Email: asimpson@ludwig.org.br
          This sequence was derived from the FAPESP/LICR Human Cancer Genome
          Project. This entry can be seen in the following URL
          (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=PM0-AN0087-130
          900-003-all&t3=2000-09-13&t4=1)
          Seq primer: puc 18 forward
          High quality sequence start: 12
          High quality sequence stop: 407.
FEATURES  Location/Qualifiers
            source
              1..407
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="AN0087"
                /dev_stage="Adult"
                /note="Organ: amnion_normal; Vector: puc18; Site_1: SmaI;
                Site_2: SmaI; A mini-library was made by cloning products
                derived from ORESTES PCR (U.S. Letters Patent application
                No. 196,716 - Ludwig Institute for Cancer Research)
                profiles into the pUC 18 vector. Reverse transcription of
                tissue mRNA and cDNA amplification were performed under
                low stringency conditions."
BASE COUNT 109 a 96 c 113 g 88 t 1 others
ORIGIN
          1 tcgtcaggtgcaggtcagcagcttg 25
          2 cgtcaggtgcaggtcagcagcttg 25
          Db 146 CTTCAGGTGCAGGTAGTACGTGG 123

RESULT 17
AW837781
LOCUS     AW837781 429 bp mRNA EST 18-MAY-2000
DEFINITION CMI-LF0042-100300-140-b12 LT0042 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW837781
VERSION   AW837781.1 GI:7931755
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
           1 (bases 1 to 429)
           Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
           Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
           Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
           Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
           ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
           Simpson,A.J.
TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
          sequence tags
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

REFERENCE 1 (bases 1 to 407)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
          Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
          Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
          Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
          ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
          Simpson,A.J.
TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
          sequence tags
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

```

```

MEDLINE 20202663
COMMENT  Contact: Simpson A.J.G.
        Laboratory of Cancer Genetics
        Ludwig Institute for Cancer Research
        Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
        Brazil
        Tel: +55-11-2704922
        Fax: +55-11-2707001
        Email: asimpson@ludwig.org.br
        This sequence was derived from the FAPESP/LICR Human Cancer Genome
        Project. This entry can be seen in the following URL
        (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=CM1-LT0042-100
        300-140-b12&t3=2000-03-10&t4=1)
        Seq primer: puc 18 forward
        High quality sequence start: 9
        High quality sequence stop: 429.
FEATURES  Location/Qualifiers
            source
              1..429
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="LT0042"
                /dev_stage="Adult"
                /note="Organ: leiomyos; Vector: puc18; Site_1: SmaI;
                Site_2: SmaI; A mini-library was made by cloning products
                derived from ORESTES PCR (U.S. Letters Patent application
                No. 196,716 - Ludwig Institute for Cancer Research)
                profiles into the pUC 18 vector. Reverse transcription of
                tissue mRNA and cDNA amplification were performed under
                low stringency conditions."
BASE COUNT 73 a 132 c 92 g 132 t
ORIGIN
          1 tcgtcaggtgcaggtcagcagctt 24
          2 tcgtcaggtgcaggtcagcagctt 24
          Db 58 TCTTCAGGTGCAGGTCTGGACCTT 81

RESULT 18
AW441001/c
LOCUS     AW441001 460 bp DNA GSS 03-OCT-2000
DEFINITION LM0232624F Mouse 10kb plasmid UUC1M library Mus musculus genomic
          clone UUC1M0232624 F, DNA sequence.
ACCESSION AW441001
VERSION   AW441001.1 GI:10565014
KEYWORDS  GSS.
SOURCE    house mouse.
ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
           1 (bases 1 to 460)
           Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
           Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly
           ,M., Rose,M., Rose,R., Stokes,R., Stokes,R., Tingey,A., von Niederhausen,A.
           and Wright,D., Weiss,R.
           Mouse whole genome scaffolding with paired end reads from 10kb
           plasmid inserts
           Unpublished (2000)
           Contact: Robert B. Weiss
           University of Utah Genome Center
           University of Utah
           Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
           84112, USA
           Tel: 801 585 5606
           Fax: 801 585 7177
           Email: ddunn@genetics.utah.edu
           Insert Length: 10000 Std Error: 0.00
           Plate: 0232 row: G column: 24
           Seq primer: CGTTGTAACCGACGGCAGT

```

OS	<i>Hordeum vulgare</i> (barley)
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC	Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
XX	
RN	[1]
RP	1-513
RA	Wing R., Close T.J., Kleinhofs A., Wise R., Begum D., Frisch D., Yu Y.,
RA	Anderson H., Dale J., Henry D., Karnodle S., Palmer M., Rambo T.,
RA	Saski C., Schwartzbeck J., Simmons J., Choi D.W., Main D., Wood T.;
RT	"Development of a genetically and physically anchored EST resource for
RT	barley genomics";
RL	Unpublished.
XX	

High quality sequence stop: 574.

# FEATURES

Location/Qualifiers  
1. .575  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="AN0087"  
/dev\_stage="Adult"  
/note="Organ: amnion normal; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 159 a 146 c 157 g 112 t 1 others  
ORIGIN

Query Match 70.4%; Score 17.6; DB 168; Length 575;  
Best Local Similarity 83.3%; Pred. No. 9e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cgtcagggtcaggtcagcagcttg 25

Db 135 CTTGAGGTGTCAGGTGAGTACGTGG 112

# RESULT 21

LOCUS BF233227/c 871 bp mRNA EST 14-NOV-2000  
DEFINITION 602023781F1 NCI\_CGAP\_L19 Mus musculus cDNA clone IMAGE:4159005 5',  
mRNA sequence.

ACCESSION BF233227

VERSION BF233227.1 GI:11143414

KEYWORDS EST.

SOURCE house mouse.

# ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 871)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM9436 row: a column: 22

High quality sequence stop: 626.

# FEATURES

## source

Location/Qualifiers  
1. .871  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4159005"  
/clone\_lib="NCI\_CGAP\_L19"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: liver; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.9 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

# BASE COUNT

## ORIGIN

Query Match 70.4%; Score 17.6; DB 146; Length 871;

Best Local Similarity 83.3%; Pred. No. 9.3e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 tcgtcagggtcaggtcagcagctt 24

Db 535 TCGGACAGGTAGGTGTCAGGTGCTT 512

# RESULT 22

## CNS02KFE/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## DEFINITION

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1077)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@email.nih.gov
Tissue Procurement: DCTD/DTP/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9661 row: d column: 10
High quality sequence stop: 643.
FEATURES
    source
    1..1077
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:3885657"
        /clone_lib="NIH_MGC_69"
        /tissue_type="large cell carcinoma, undifferentiated"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
        Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
        Average insert size 1.1 kb. Library constructed by Life
        Technologies."
BASE COUNT 324 a 278 c 322 g 153 t
ORIGIN
Query Match 70.4%; Score 17.6; DB 138; Length 1077;
Best Local Similarity 83.3%; Pred. No. 9.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 cgtcaggtcagggtcagcagcttg 25
| ||||| ||||| || |||||
Db 228 CTTGAGGTGCGAGTAAGTACGTGG 205

RESULT 24
LOCUS CNS03FBT/1079 bp DNA GSS 17-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
021L24 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL241490.1 GI:7962259
VERSION GSS: genome survey sequence.
KEYWORDS Tetraodon nigroviridis
SOURCE Tetraodon nigroviridis
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 1079)
AUTHORS Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1079)
AUTHORS Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1079)

```

---

```

Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
    Location/Qualifiers
    1..1079
        /organism="Tetraodon nigroviridis"
        /db_xref="taxon:99883"
        /clone="021L24"
        /clone_lib="G"
        /note="Genoscope sequence ID : COBG021DF12SP1-end ;
        PUC-ori"
BASE COUNT 263 a 247 c 279 g 288 t 2 others
ORIGIN
Query Match 70.4%; Score 17.6; DB 221; Length 1079;
Best Local Similarity 83.3%; Pred. No. 9.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 tgcgcaggtcagggtcagcagctt 24
| ||||| ||||| || |||||
Db 1036 TCCTGAGGTGAAAGTCAGCACTTT 1013

RESULT 25
LOCUS BB516544/288 bp mRNA EST 28-JUL-2000
DEFINITION musculus cDNA clone D830016J11 3' similar to M76601 Mouse alpha
cardiac myosin heavy chain mRNA, mRNA sequence.
ACCESSION BB516544.1 GI:9568002
VERSION BB516544.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 288)
AUTHORS Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata
Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya
T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.,
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino
M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp
URL: http://genome.rtc.riken.go.jp/

```

Sequence Tagged Connector  
Plate: 3110 row: F column: 23  
Class: BAC ends  
High quality sequence stop: 398.

FEATURES	
SOURCE	

Query Match	69.6%;	Score 17.4;	DB 226;	Length 398;
Best Local Similarity	94.7%;	Pred. No. 1.1e+03;		
Matches 18;	Conservative	0;	Mismatches 1;	Indels 0;
				Gaps 0;

QY 7 ggtgcaggctcagcacgttg 25  
|||||  
Db 330 GGTGCAGGTCAGCACGATG 348

RESULT 27

LOCUS  
DEFINITION

ACCESSION

**VERSION**  
**KEYWORDS**

REFERENCE  
AUTHORS

100

TITLE

JOURNAL  
MEDLINE

REFERENCE  
AUTHORS

1  
2  
3  
4  
5

TITLE	NUMBER
1. The first part of the report is a general introduction to the project.	1
2. The second part of the report is a detailed description of the methodology used.	2
3. The third part of the report is a presentation of the results of the study.	3
4. The fourth part of the report is a discussion of the findings and their implications.	4
5. The fifth part of the report is a conclusion and a list of references.	5

JOURNAL  
MEDLINE  
REFERENCE

REFERENCE	AUTHORS	TITLE
1	...	...
2	...	...
3	...	...
4	...	...
5	...	...
6	...	...
7	...	...
8	...	...
9	...	...
10	...	...
11	...	...
12	...	...
13	...	...
14	...	...
15	...	...
16	...	...
17	...	...
18	...	...
19	...	...
20	...	...
21	...	...
22	...	...
23	...	...
24	...	...
25	...	...
26	...	...
27	...	...
28	...	...
29	...	...
30	...	...
31	...	...
32	...	...
33	...	...
34	...	...
35	...	...
36	...	...
37	...	...
38	...	...
39	...	...
40	...	...
41	...	...
42	...	...
43	...	...
44	...	...
45	...	...
46	...	...
47	...	...
48	...	...
49	...	...
50	...	...
51	...	...
52	...	...
53	...	...
54	...	...
55	...	...
56	...	...
57	...	...
58	...	...
59	...	...
60	...	...
61	...	...
62	...	...
63	...	...
64	...	...
65	...	...
66	...	...
67	...	...
68	...	...
69	...	...
70	...	...
71	...	...
72	...	...
73	...	...
74	...	...
75	...	...
76	...	...
77	...	...
78	...	...
79	...	...
80	...	...
81	...	...
82	...	...
83	...	...
84	...	...
85	...	...
86	...	...
87	...	...
88	...	...
89	...	...
90	...	...
91	...	...
92	...	...
93	...	...
94	...	...
95	...	...
96	...	...
97	...	...
98	...	...
99	...	...
100	...	...

JOURNAL  
COMMENT

## FEATURES

source

PAGE COUNT

BASE COUNT

1



## ORIGIN

Query Match 69.6%; Score 17.4; DB 222; Length 1101;  
 Best Local Similarity 94.7%; Pred. No. 1.2e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 tcaggtgcaggtcagcagc 22  
 |||||  
 Db 759 TCAGGTGCAGGTCCACACG 741

## RESULT 28

BG561755 251 bp mRNA EST 10-APR-2001  
 LOCUS ETESTed99h03.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria  
 tenella cDNA 5', mRNA sequence.

ACCESSION BG561755  
 VERSION BG561755.1 GI:13590753

KEYWORDS EST.  
 SOURCE Eimeria tenella.  
 ORGANISM Eimeria tenella.  
 Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;  
 Eimeria.

## REFERENCE

1 (bases 1 to 251)  
 Liberato P., Diaz, C., Tang, K., Marra, M., Hillier, L., Kucaba, T.,  
 Martin, J., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen  
 M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey  
 N., Schurk, R., Ritter, E., Kohn, S., Florence, N., Shin, T., Jackson  
 Y., Cardenas, M., McCann, R., Waterston, R., Wilson, R. and Sibley, D.  
 WashU-Merck Eimeria tenella project  
 Unpublished (1999)

## TITLE

WashU-Merck Eimeria tenella project

## JOURNAL

WashU-Merck Eimeria tenella project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Contact David Sibley (toxest@borcim.wustl.edu) for further

information relating to organism, libraries, or clone availability.

Seq primer: -40RP from Gibco.

## FEATURES

Source

1..251  
 /organism="Eimeria tenella"  
 /strain="LS18"  
 /db\_xref="taxon:5802"  
 /clone\_lib="Eimeria tenella S5-2 cDNA Neg Selected"  
 /dev\_stage="Sporozoite stage"  
 /lab\_host="SOLR"  
 /note="Vector: Bluescript SK-; Site.1: EcoRI; Site.2: XhoI  
 ; Sporozoites were obtained from in vitro sporulated and  
 excysted oocysts of E. tenella grown in chickens. cDNA was  
 synthesized from poly mRNA using an oligo-dT primer  
 containing a XhoI site. Following second strand synthesis,  
 EcoRI adapters were ligated to the cDNA and products were  
 size-selected on sephacryl S500. The cDNA were ligated to  
 EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were  
 converted to phagemids by mass excision using EXASSIST  
 helper phage and E.coli SOLR cell (Stratagene). Clones  
 were selected by negative hybridization against a pool of  
 overrepresented ESTs (N=10, from 1682 previous reads).  
 Insert sizes range from 1.2-2.9Kb. The library may contain  
 a small percentage of host or bacterial contaminants."  
 59 a 52 c 67 g 73 t

BASE COUNT  
 ORIGIN

Query Match 68.8%; Score 17.2; DB 155; Length 251;  
 Best Local Similarity 86.4%; Pred. No. 1.3e+03;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 tcaggtgcaggtcagcagcgttg 25

Db 141 TCATGTGCAGGTTCATCATTTG 162  
 ||| ||||| ||||| ||||| |||||

## RESULT 29

AI755334/c

LOCUS

DEFINITION

5', mRNA sequence.

ACCESSION AI755334

VERSION AI755334.1 GI:5149133

KEYWORDS EST.

SOURCE

ORGANISM

Eimeria tenella.

Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;  
 Eimeria.

REFERENCE

AUTHORS

1 (bases 1 to 303)

Liberato P., Diaz, C., Tang, K., Marra, M., Hillier, L., Kucaba, T.,  
 Martin, J., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen  
 M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey  
 N., Schurk, R., Ritter, E., Kohn, S., Florence, N., Shin, T., Jackson  
 Y., Cardenas, M., McCann, R., Waterston, R., Wilson, R. and Sibley, D.  
 WashU-Merck Eimeria tenella project  
 Unpublished (1999)

CONTACT: David Sibley, Ph.D.

WashU-Merck Eimeria tenella project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Contact David Sibley (toxest@borcim.wustl.edu) for further

information relating to organism, libraries, or clone availability.

Seq primer: -40RP from Gibco

High quality sequence stop: 280.

FEATURES

source

1..303

/organism="Eimeria tenella"

/strain="LS18"

/db\_xref="taxon:5802"

/clone\_lib="Eimeria S5-2 Sporozoite stage"

/dev\_stage="Sporozoite"

/lab\_host="SOLR E. coli"

/note="Vector: Bluescript SK-; Site.1: EcoRI; Site.2: XhoI  
 ; Sporozoites were obtained from in vitro sporulated and  
 excysted oocysts of E. tenella grown in chickens. cDNA  
 was synthesized from poly mRNA using an oligo-dT primer  
 containing a XhoI site. Following second strand synthesis,  
 EcoRI adapters were ligated to the cDNA and products were  
 size-selected on Sephadryl S500. cDNAs were digested with  
 EcoRI/XhoI and cloned into lambda Zap II (Stratagene).  
 Clones were converted to phagemids by mass excision using  
 EXASSIST helper phage and SOLR cells (Stratagene).  
 Insert sizes range from 1.2-2.9 kb."  
 87 a 80 c 58 g 77 t 1 others

BASE COUNT  
 ORIGIN

Query Match 68.8%; Score 17.2; DB 24; Length 303;  
 Best Local Similarity 86.4%; Pred. No. 1.3e+03;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 tcaggtgcaggtcagcagcgttg 25

|||||

Db 129 TCATGTGCAGGTTCATCATTTG 108

## RESULT 30

BF331947/c

LOCUS

DEFINITION

QV2-BT0634-280800-331-e08 BT0634 Homo sapiens cDNA, mRNA sequence.

ACCESSION BF331947

VERSION BF331947.1 GI:11302695

BF331947 379 bp mRNA EST

18-JAN-2000



Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 Contact David Sibley (toxoe@borcim.wustl.edu) for further  
 information relating to organism, libraries, or clone availability.  
 Seq primer: -40RP from Gbco.

## FEATURES

source

```

Location/Qualifiers
1..408
/organism="Eimeria tenella"
/strain="LS18"
/db_xref="taxon:5802"
/clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
/dev_stage="Sporozoite stage"
/lab_host="SOLR"
/Note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
; Sporozoites were obtained from in vitro sporulated and
excysted oocysts of E. tenella grown in chickens. cDNA was
synthesized from poly mRNA using an oligo-dT primer
containing a XhoI site. Following second strand synthesis,
EcoRI adapters were ligated to the cDNA and products were
size-selected on sephacryl S500. The cDNA were ligated to
EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
converted to phagemids by mass excision using ExAssist
helper phage and E.coli SOLR cell (Stratagene). Clones
were selected by negative hybridization against a pool of
overrepresented ESTs (N>=10, from 1682 previous reads).
Insert sizes range from 1.2-2.9Kb. The library may contain
a small percentage of host or bacterial contaminants."

BASE COUNT      99 a  77 c  114 g  118 t
ORIGIN
Query Match      68.8%; Score 17.2; DB 154; Length 408;
Best Local Similarity 86.4%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  4  tcagggtgcaggtcagcagcttg 25
      ||| ||||| ||||| ||| |||
Db   304  TCATGTGCAGGTCATCATG 325

RESULT 33
BG466711
LOCUS      408 bp  mRNA      EST      20-MAR-2001
DEFINITION ETESTed37c04.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
tenella cDNA 5', mRNA sequence.
ACCESSION  BG466711
VERSION     BG466711.1 GI:13395686
KEYWORDS   Eimeria tenella.
SOURCE     Eimeria tenella
ORGANISM   Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
Eimeria.
REFERENCE  1 (bases 1 to 408)
AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
WashU-Merck Eimeria tenella project
Unpublished (1999)
Contact: David Sibley, Ph.D.
WashU-Merck Eimeria tenella project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Contact David Sibley (toxoe@borcim.wustl.edu) for further
information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gbco.
Location/Qualifiers
1..408
/organism="Eimeria tenella"

```

/strain="LS18"  
/db\_xref="taxon:5802"  
/clone\_lib="Eimeria tenella S5-2 cDNA Neg Selected"  
/dev\_stage="Sporozoite stage"  
/lab\_host="SOLR"  
/Note="Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI
; Sporozoites were obtained from in vitro sporulated and
excysted oocysts of E. tenella grown in chickens. cDNA was
synthesized from poly mRNA using an oligo-dT primer
containing a XhoI site. Following second strand synthesis,
EcoRI adapters were ligated to the cDNA and products were
size-selected on sephacryl S500. The cDNA were ligated to
EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
converted to phagemids by mass excision using ExAssist
helper phage and E.coli SOLR cell (Stratagene). Clones
were selected by negative hybridization against a pool of
overrepresented ESTs (N>=10, from 1682 previous reads).
Insert sizes range from 1.2-2.9Kb. The library may contain
a small percentage of host or bacterial contaminants."

BASE COUNT 99 a 77 c 114 g 118 t  
ORIGIN

Query Match 68.8%; Score 17.2; DB 154; Length 408;  
Best Local Similarity 86.4%; Pred. No. 1.3e+03;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 tcagggtgcaggtcagcagcttg 25  
 ||| ||||| ||||| ||| |||  
Db 304 TCATGTGCAGGTCATCATG 325

RESULT 34  
AI759462/c  
LOCUS 410 bp mRNA EST 18-JAN-2000  
DEFINITION ETEStea22a07.x1 Eimeria S5-2 Sporozoite stage Eimeria tenella cDNA  
3', mRNA sequence.  
ACCESSION AI759462  
VERSION AI759462.1 GI:5175213  
KEYWORDS EST.  
SOURCE Eimeria tenella.  
ORGANISM Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;  
Eimeria.

REFERENCE 1 (bases 1 to 410)  
AUTHORS Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,  
Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen  
,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey  
,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson  
,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.  
WashU-Merck Eimeria tenella project  
Unpublished (1999)  
Other\_ESTs: ea22a07.y1  
Contact: David Sibley, Ph.D.  
WashU-Merck Eimeria tenella project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Contact David Sibley (toxoe@borcim.wustl.edu) for further  
information relating to organism, libraries, or clone availability.  
Seq primer: -40UP from Gbco  
High quality sequence stop: 376.  
Location/Qualifiers  
1..410  
/organism="Eimeria tenella"  
/strain="LS18"  
/db\_xref="taxon:5802"  
/clone\_lib="Eimeria S5-2 Sporozoite stage"  
/dev\_stage="Sporozoite"  
/lab\_host="SOLR E. coli"  
/Note="Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI

## FEATURES

source

```

Location/Qualifiers
1..410
/organism="Eimeria tenella"
/strain="LS18"
/db_xref="taxon:5802"
/clone_lib="Eimeria S5-2 Sporozoite stage"
/dev_stage="Sporozoite"
/lab_host="SOLR E. coli"
/Note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI

```



```

Query Match      68.8%; Score 17.2; DB 154; Length 414;
Best Local Similarity 86.4%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 tcaggtgcaggtcagcagcttg 25
   ||| ||||| ||||| ||| |||
Db 140 TCATGTGCAGGTCATCATTG 119

RESULT 37
AI755467/c
LOCUS      AI755467      418 bp      mRNA      EST      18-JAN-2000
DEFINITION ETESte37c06.y1 Eimeria S5-2 Sporozoite stage Eimeria tenella cDNA
5' mRNA sequence.
ACCESSION  AI755467
VERSION     AI755467.1 GI:5149190
KEYWORDS   Eimeria tenella.
SOURCE     Eimeria tenella.
ORGANISM   Eimeria tenella.
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;

REFERENCE  1 (bases 1 to 418)
AUTHORS   Liberators P., Diaz, C., Tang, K., Marra, M., Hillier, L., Kucaba, T.,
            Martin, J., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen
            M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey
            N., Schurk, R., Ritter, E., Kohn, S., Florence, N., Shin, T., Jackson
            Y., Cardenas, M., McCann, R., Waterston, R., Wilson, R. and Sibley, D.
            WashU-Merck Eimeria tenella project
            Unpublished (1999)
            Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 411.
            Location/Qualifiers
                1..418
                /organism="Eimeria tenella"
                /strain="LS18"
                /db_xref="taxon:5802"
                /clone_lib="Eimeria S5-2 Sporozoite stage"
                /dev_stage="Sporozoite"
                /lab_host="SOLR E. coli"
                /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                ; Sporozoites were obtained from in vitro sporulated and
                excysted oocysts of E. tenella grown in chickens. cDNA
                was synthesized from poly mRNA using an oligo-dT primer
                containing a XhoI site. Following second strand synthesis,
                EcoRI adapters were ligated to the cDNA and products were
                size-selected on Sephacryl S500. cDNAs were digested with
                EcoRI/XhoI and cloned into lambda Zap II (Stratagene).
                Clones were converted to phagemids by mass excision using
                ExAssist helper phage and SOLR cells (Stratagene).
                Insert sizes range from 1.2-2.9 kb."
                Insert sizes range from 1.2-2.9 kb."

BASE COUNT  119 a 115 c 74 g 110 t
ORIGIN

Query Match      68.8%; Score 17.2; DB 24; Length 418;
Best Local Similarity 86.4%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 tcaggtgcaggtcagcagcttg 25
   ||| ||||| ||||| ||| |||
Db 139 TCATGTGCAGGTCATCATTG 118

RESULT 38
AI755467/c
LOCUS      AI755467      418 bp      mRNA      EST      10-APR-2001
DEFINITION ETESte01e01.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
tenella cDNA 5', mRNA sequence.
ACCESSION  BG561583
VERSION     BG561583.1 GI:13590581
KEYWORDS   Eimeria tenella.
SOURCE     Eimeria tenella.
ORGANISM   Eimeria tenella.
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;

REFERENCE  1 (bases 1 to 418)
AUTHORS   Liberators P., Diaz, C., Tang, K., Marra, M., Hillier, L., Kucaba, T.,
            Martin, J., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen
            M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey
            N., Schurk, R., Ritter, E., Kohn, S., Florence, N., Shin, T., Jackson
            Y., Cardenas, M., McCann, R., Waterston, R., Wilson, R. and Sibley, D.
            WashU-Merck Eimeria tenella project
            Unpublished (1999)
            Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 417.
            Location/Qualifiers
                1..418
                /organism="Eimeria tenella"
                /strain="LS18"
                /db_xref="taxon:5802"
                /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
                /dev_stage="Sporozoite stage"
                /lab_host="SOLR"
                /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                ; Sporozoites were obtained from in vitro sporulated and
                excysted oocysts of E. tenella grown in chickens. cDNA
                was synthesized from poly mRNA using an oligo-dT primer
                containing a XhoI site. Following second strand synthesis,
                EcoRI adapters were ligated to the cDNA and products were
                size-selected on Sephacryl S500. The cDNA were ligated to
                EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
                converted to phagemids by mass excision using ExAssist
                helper phage and E. coli SOLR cell (Stratagene). Clones
                were selected by negative hybridization against a pool of
                overrepresented ESTs (N>=10, from 1682 previous reads).
                Insert sizes range from 1.2-2.9Kb. The library may contain
                a small percentage of host or bacterial contaminants."
                Insert sizes range from 1.2-2.9Kb. The library may contain
                a small percentage of host or bacterial contaminants."

BASE COUNT  118 a 113 c 73 g 112 t
ORIGIN

Query Match      68.8%; Score 17.2; DB 154; Length 418;
Best Local Similarity 86.4%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 tcaggtgcaggtcagcagcttg 25
   ||| ||||| ||||| ||| |||
Db 142 TCATGTGCAGGTCATCATTG 121

RESULT 39
BG561583/c
LOCUS      BG561583      421 bp      mRNA      EST      10-APR-2001
DEFINITION ETESte01e01.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
tenella cDNA 5', mRNA sequence.
ACCESSION  BG561583
VERSION     BG561583.1 GI:13590581
KEYWORDS   Eimeria tenella.
SOURCE     Eimeria tenella.
ORGANISM   Eimeria tenella.
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;

REFERENCE  1 (bases 1 to 418)
AUTHORS   Liberators P., Diaz, C., Tang, K., Marra, M., Hillier, L., Kucaba, T.,
            Martin, J., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen
            M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey
            N., Schurk, R., Ritter, E., Kohn, S., Florence, N., Shin, T., Jackson
            Y., Cardenas, M., McCann, R., Waterston, R., Wilson, R. and Sibley, D.
            WashU-Merck Eimeria tenella project
            Unpublished (1999)
            Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 417.
            Location/Qualifiers
                1..418
                /organism="Eimeria tenella"
                /strain="LS18"
                /db_xref="taxon:5802"
                /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
                /dev_stage="Sporozoite stage"
                /lab_host="SOLR"
                /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                ; Sporozoites were obtained from in vitro sporulated and
                excysted oocysts of E. tenella grown in chickens. cDNA
                was synthesized from poly mRNA using an oligo-dT primer
                containing a XhoI site. Following second strand synthesis,
                EcoRI adapters were ligated to the cDNA and products were
                size-selected on Sephacryl S500. The cDNA were ligated to
                EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
                converted to phagemids by mass excision using ExAssist
                helper phage and E. coli SOLR cell (Stratagene). Clones
                were selected by negative hybridization against a pool of
                overrepresented ESTs (N>=10, from 1682 previous reads).
                Insert sizes range from 1.2-2.9Kb. The library may contain
                a small percentage of host or bacterial contaminants."
                Insert sizes range from 1.2-2.9Kb. The library may contain
                a small percentage of host or bacterial contaminants."

BASE COUNT  118 a 113 c 73 g 112 t
ORIGIN

```

SOURCE  
ORGANISM  
Eimeria tenella.  
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;  
Eimeria.

REFERENCE  
1 (bases 1 to 421)

AUTHORS  
Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,  
Martin,J., Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen  
M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey  
N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson  
Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.  
WashU-Merck Eimeria tenella project  
Unpublished (1999)

TITLE  
WashU-Merck Eimeria tenella project

JOURNAL  
Contact: David Sibley, Ph.D.  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

COMMENT  
Contact David Sibley (toxoe@borcim.wustl.edu) for further  
information relating to organism, libraries, or clone availability.  
Seq primer: -40RP from Gibco  
High quality sequence stop: 400.

FEATURES  
Location/Qualifiers  
source  
1..421  
/organism="Eimeria tenella"  
/strain="LS18"  
/db\_xref="taxon:5802"  
/clone\_lib="Eimeria tenella S5-2 cDNA Neg Selected"  
/dev\_stage="Sporozoite stage"  
/lab\_host="SOLR"  
/note="Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI  
; Sporozoites were obtained from in vitro sporulated and  
excysted oocysts of E. tenella grown in chickens. cDNA was  
synthesized from poly mRNA using an oligo-dT primer  
containing a XhoI site. Following second strand synthesis,  
EcoRI adapters were ligated to the cDNA and products were  
size-selected on sephacryl S500. The cDNA were ligated to  
EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were  
converted to phagemids by mass excision using ExAssist  
helper phage and E.coli SOLR cell (Stratagene). Clones  
were selected by negative hybridization against a pool of  
overrepresented ESTs (N>=10, from 1682 previous reads).  
Insert sizes range from 1.2-2.9Kb. The library may contain  
a small percentage of host or bacterial contaminants."

BASE COUNT  
116 a 112 c 74 g 119 t

ORIGIN

Query Match 68.8%; Score 17.2; DB 155; Length 421;  
Best Local Similarity 86.4%; Pred. No. 1.3e+03;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 tcaggtgcaggtcagcagcttg 25  
|||||  
Db 153 TCATGTCAGGTCATCACATTG 132

RESULT 40  
A1757999/c

LOCUS  
DEFINITION  
A1757999 427 bp mRNA EST 18-JAN-2000  
ETESTea35h05.y1 Eimeria S5-2 sporozoite stage Eimeria tenella cDNA  
5', mRNA sequence.

ACCESSION  
A1757999

VERSION  
A1757999.1 GI:5151722

KEYWORDS  
EST.

SOURCE  
Eimeria tenella.  
Eimeria tenella  
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;  
Eimeria.

REFERENCE  
1 (bases 1 to 427)

AUTHORS  
Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,  
Martin,J., Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen  
M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey  
N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson  
Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.  
WashU-Merck Eimeria tenella project  
Unpublished (1999)

TITLE  
WashU-Merck Eimeria tenella project

JOURNAL  
Contact: David Sibley, Ph.D.  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

COMMENT  
Contact David Sibley (toxoe@borcim.wustl.edu) for further  
information relating to organism, libraries, or clone availability.  
Seq primer: -40RP from Gibco  
High quality sequence stop: 392.

FEATURES  
Location/Qualifiers  
source  
1..427  
/organism="Eimeria tenella"  
/strain="LS18"  
/db\_xref="taxon:5802"  
/clone\_lib="Eimeria S5-2 Sporozoite stage"  
/dev\_stage="Sporozoite"  
/lab\_host="SOLR E. coli"  
/note="Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI  
; Sporozoites were obtained from in vitro sporulated and  
excysted oocysts of E. tenella grown in chickens. cDNA  
was synthesized from poly mRNA using an oligo-dT primer  
containing a XhoI site. Following second strand synthesis,  
EcoRI adapters were ligated to the cDNA and products were  
size-selected on Sephacryl S500. cDNAs were digested with  
EcoRI/XhoI and cloned into lambda Zap II (Stratagene).  
Clones were converted to phagemids by mass excision using  
ExAssist helper phage and SOLR cells (Stratagene).  
Insert sizes range from 1.2-2.9 kb."

BASE COUNT  
125 a 113 c 74 g 115 t

ORIGIN

Query Match 68.8%; Score 17.2; DB 24; Length 427;  
Best Local Similarity 86.4%; Pred. No. 1.3e+03;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 tcaggtgcaggtcagcagcttg 25  
|||||  
Db 133 TCATGTCAGGTCATCACATTG 112

RESULT 41  
BG235712/c

LOCUS  
DEFINITION  
BG235712 442 bp mRNA EST 12-FEB-2001  
ETESTdlhl1.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria  
tenella cDNA 5', mRNA sequence.

ACCESSION  
BG235712

VERSION  
BG235712.1 GI:12749559

KEYWORDS  
EST.

SOURCE  
Eimeria tenella.  
Eimeria tenella  
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;  
Eimeria.

REFERENCE  
1 (bases 1 to 442)

AUTHORS  
Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,  
Martin,J., Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen  
M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey  
N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson  
Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.  
WashU-Merck Eimeria tenella project  
Unpublished (1999)

TITLE  
WashU-Merck Eimeria tenella project

JOURNAL  
Contact: David Sibley, Ph.D.  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800

M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey  
N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson  
Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.  
WashU-Merck Eimeria tenella project  
Unpublished (1999)

TITLE  
WashU-Merck Eimeria tenella project

JOURNAL  
Contact: David Sibley, Ph.D.  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

COMMENT  
Contact David Sibley (toxoe@borcim.wustl.edu) for further  
information relating to organism, libraries, or clone availability.  
Seq primer: -40RP from Gibco  
High quality sequence stop: 392.

FEATURES  
Location/Qualifiers  
source  
1..427  
/organism="Eimeria tenella"  
/strain="LS18"  
/db\_xref="taxon:5802"  
/clone\_lib="Eimeria S5-2 Sporozoite stage"  
/dev\_stage="Sporozoite"  
/lab\_host="SOLR E. coli"  
/note="Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI  
; Sporozoites were obtained from in vitro sporulated and  
excysted oocysts of E. tenella grown in chickens. cDNA  
was synthesized from poly mRNA using an oligo-dT primer  
containing a XhoI site. Following second strand synthesis,  
EcoRI adapters were ligated to the cDNA and products were  
size-selected on Sephacryl S500. cDNAs were digested with  
EcoRI/XhoI and cloned into lambda Zap II (Stratagene).  
Clones were converted to phagemids by mass excision using  
ExAssist helper phage and SOLR cells (Stratagene).  
Insert sizes range from 1.2-2.9 kb."

BASE COUNT  
125 a 113 c 74 g 115 t

ORIGIN

Query Match 68.8%; Score 17.2; DB 24; Length 427;  
Best Local Similarity 86.4%; Pred. No. 1.3e+03;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 tcaggtgcaggtcagcagcttg 25  
|||||  
Db 133 TCATGTCAGGTCATCACATTG 112

RESULT 41  
BG235712/c

LOCUS  
DEFINITION  
BG235712 442 bp mRNA EST 12-FEB-2001  
ETESTdlhl1.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria  
tenella cDNA 5', mRNA sequence.

ACCESSION  
BG235712

VERSION  
BG235712.1 GI:12749559

KEYWORDS  
EST.

SOURCE  
Eimeria tenella.  
Eimeria tenella  
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;  
Eimeria.

REFERENCE  
1 (bases 1 to 442)

AUTHORS  
Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,  
Martin,J., Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen  
M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey  
N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson  
Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.  
WashU-Merck Eimeria tenella project  
Unpublished (1999)

TITLE  
WashU-Merck Eimeria tenella project

JOURNAL  
Contact: David Sibley, Ph.D.  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800

Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 Contact David Sibley (toxoe@borcim.wustl.edu) for further  
 information relating to organism, libraries, or clone availability.  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 422.  
 Location/Qualifiers

## FEATURES

source

1. .442  
 /organism="Eimeria tenella"  
 /strain="LS18"  
 /db\_xref="taxon:5802"  
 /clone\_lib="Eimeria tenella S5-2 cDNA Neg Selected"  
 /dev\_stage="Sporozoite stage"  
 /lab\_host="SOLR"  
 /note="Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI  
 ; Sporozoites were obtained from in vitro sporulated and  
 excysted oocysts of E. tenella grown in chickens. cDNA was  
 synthesized from poly mRNA using an oligo-dT primer  
 containing a XhoI site. Following second strand synthesis,  
 EcoRI adapters were ligated to the cDNA and products were  
 size-selected on sephacryl S500. The cDNA were ligated to  
 EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were  
 converted to phagemids by mass excision using ExAssist  
 helper phage and E.coli SOLR cell (Stratagene). Clones  
 were selected by negative hybridization against a pool of  
 overrepresented ESTs (N>=10, from 1682 previous reads).  
 Insert sizes range from 1.2-2.9Kb. The library may contain  
 a small percentage of host or bacterial contaminants."  
 a 118 C 78 g 117 t

BASE COUNT  
 ORIGIN

Query Match 68.8%; Score 17.2; DB 175; Length 442;  
 Best Local Similarity 86.4%; Pred. No. 1.3e+03;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 4 tcaggtgcaggtcagcagcttg 25  
 ||| ||||| ||||| ||| |||  
 Db 131 TCATGTGCGAGGTGCATCATTTG 110

RESULT 42

A0444310

LOCUS

DEFINITION GSSTc0372 Trypanosoma cruzi random genomic library Trypanosoma  
 cruzi genomic clone G1H13, DNA sequence.

ACCESSION

A0444310

VERSION

A0444310.3

KEYWORDS

GSS.

SOURCE

Trypanosoma cruzi.

ORGANISM

Trypanosoma cruzi

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

1 (bases 1 to 443)

Aguero, F., Verdun, R., Frasch, A.C.C. and Sanchez, D.O.

A random sequencing approach for the analysis of the trypanosoma

cruzi genome: general structure, large gene and repetitive DNA

families, and gene discovery

Genome Res. 10 (12), 1996-2005 (2000)

20568489

On Sep 14, 2000 this sequence version replaced gi:9374071.

Contact: Sanchez D.O.

Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral

San Martin)

Av. Gral Paz entre Albarillos y Constituyentes, INTI edificio 24

CP(1650) San Martin, Prov. de BS AS. Argentina

Tel: 54-11-4580-7255 ext 309

Fax: 54-11-4752-9639

Email: dsanchez@lib.unsam.edu.ar

Sequences were basecalled with phred and vector was masked with

crossmatch (see http://genome.washington.edu). Sequences were then

trimmed from both ends to remove low quality bases and masked

vector.

Seq primer: T7  
 Class: shotgun.

## FEATURES

source

1. .443  
 /organism="Trypanosoma cruzi"  
 /strain="CL-Brener"  
 /db\_xref="taxon:5693"  
 /clone\_lib="Trypanosoma cruzi random genomic library"  
 /cell\_type="epimastigote"  
 /note="Vector: pBS(-) (Stratagene); T. cruzi DNA was  
 randomly sheared using a nebulizer and the 1 to 2 Kb range  
 was gel purified and cloned into the dephosphorylated  
 HindII site of the vector"  
 a 103 c 131 g 98 t

BASE COUNT  
 ORIGIN

Query Match 68.8%; Score 17.2; DB 228; Length 443;  
 Best Local Similarity 86.4%; Pred. No. 1.3e+03;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 tcgtcaggtgcaggtcagcagc 22  
 ||||| ||||| ||||| ||| |||  
 Db 29 TCGTCACGTGCGAGGCAGAACG 50

RESULT 43  
 BG516136/c

LOCUS

DEFINITION

EteSTed49g01.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria

tenella cDNA 5', mRNA sequence.

ACCESSION

BG516136

VERSION

BG516136.1

KEYWORDS

EST.

SOURCE

Eimeria tenella.

ORGANISM

Eimeria tenella

Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;

1 (bases 1 to 466)

Liberator, P., Diaz, C., Tang, K., Marra, M., Hillier, L., Kucaba, T.,

Martin, J., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen

, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey

, N., Schurk, R., Ritter, E., Kohn, S., Florence, N., Shin, T., Jackson

, Y., Cardenas, M., McCann, R., Waterston, R., Wilson, R. and Sibley, D.

WashU-Merck Eimeria tenella project

Unpublished (1999)

Contact: David Sibley, Ph.D.

WashU-Merck Eimeria tenella project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Contact David Sibley (toxoe@borcim.wustl.edu) for further

information relating to organism, libraries, or clone availability.

Seq primer: -40RP from Gibco

High quality sequence stop: 411.

Location/Qualifiers

1. .466

/organism="Eimeria tenella"

/strain="LS18"

/db\_xref="taxon:5802"

/clone\_lib="Eimeria tenella S5-2 cDNA Neg Selected"

/dev\_stage="Sporozoite stage"

/lab\_host="SOLR"

/note="Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI

; Sporozoites were obtained from in vitro sporulated and

excysted oocysts of E. tenella grown in chickens. cDNA was

synthesized from poly mRNA using an oligo-dT primer

containing a XhoI site. Following second strand synthesis,

EcoRI adapters were ligated to the cDNA and products were

size-selected on sephacryl S500. The cDNA were ligated to

ECORI/XhoI prepared lambda ZapII (Stratagene). Clones were converted to phagemids by mass excision using ExAssist helper phage and E.coli SOLR cell (Stratagene). Clones were selected by negative hybridization against a pool of overrepresented ESTs (N>=10, from 1682 previous reads). Insert sizes range from 1.2-2.9Kb. The library may contain a small percentage of host or bacterial contaminants."

BASE COUNT 133 a 125 c 85 g 123 t

ORIGIN

Query Match 68.8%; Score 17.2; DB 154; Length 466;  
Best Local Similarity 86.4%; Pred. No. 1.3e+03;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 tcaggtgcaggtcagcagcttg 25  
||| ||||| ||||| ||| |||

Db 149 TCATGTGCAGGTCAATCATTG 128

RESULT 44  
BG516408/c

LOCUS BG516408 478 bp mRNA EST 30-MAR-2001  
DEFINITION Etest58e01.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria tenella cDNA 5', mRNA sequence.

ACCESSION BG516408  
VERSION BG516408.1 GI:13489734  
KEYWORDS EST  
SOURCE Eimeria tenella.  
ORGANISM Eimeria tenella

REFERENCE 1 (bases 1 to 478)

AUTHORS Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T., Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D. WashU-Merck Eimeria tenella project  
Unpublished (1999)  
Contact: David Sibley, Ph.D.  
WashU-Merck Eimeria tenella project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

TITLE WashU-Merck Eimeria tenella project  
JOURNAL Unpublished (1999)  
COMMENT Contact: David Sibley, Ph.D.  
WashU-Merck Eimeria tenella project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Contact David Sibley (toxoe@borcim.wustl.edu) for further information relating to organism, libraries, or clone availability.  
Seq primer: -40RP from Gibco  
High quality sequence stop: 424.

#### FEATURES

source

1. 478  
/organism="Eimeria tenella"  
/strain="LS18"  
/db\_xref="taxon:5802"  
/clone\_lib="Eimeria tenella S5-2 cDNA Neg Selected"  
/dev\_stage="Sporozoite stage"  
/lab\_host="SOLR"  
/note="Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI ; Sporozoites were obtained from in vitro sporulated and excysted oocysts of E.tenella grown in chickens. cDNA was synthesized from poly mRNA using an oligo-dT primer containing a XhoI site. Following second strand synthesis, EcorI adapters were ligated to the cDNA and products were size-selected on sephacryl S500. The cDNA were ligated to EcorI/XhoI prepared lambda ZapII (Stratagene). Clones were converted to phagemids by mass excision using ExAssist helper phage and E.coli SOLR cell (Stratagene). Clones were selected by negative hybridization against a pool of overrepresented ESTs (N>=10, from 1682 previous reads). Insert sizes range from 1.2-2.9Kb. The library may contain a small percentage of host or bacterial contaminants."

BASE COUNT 136 a 126 c 88 g 128 t

ORIGIN

Query Match 68.8%; Score 17.2; DB 154; Length 478;  
Best Local Similarity 86.4%; Pred. No. 1.3e+03;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 tcaggtgcaggtcagcagcttg 25  
||| ||||| ||||| ||| |||

Db 152 TCATGTGCAGGTCAATCATTG 131

RESULT 45  
BG561248/c

LOCUS BG561248 494 bp mRNA EST 10-APR-2001  
DEFINITION Etestd82el2.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria tenella cDNA 5', mRNA sequence.

ACCESSION BG561248  
VERSION BG561248.1 GI:13590246  
KEYWORDS EST  
SOURCE Eimeria tenella.  
ORGANISM Eimeria tenella

REFERENCE 1 (bases 1 to 494)

AUTHORS Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T., Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D. WashU-Merck Eimeria tenella project  
Unpublished (1999)  
Contact: David Sibley, Ph.D.  
WashU-Merck Eimeria tenella project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

TITLE WashU-Merck Eimeria tenella project  
JOURNAL Unpublished (1999)  
COMMENT Contact: David Sibley, Ph.D.  
WashU-Merck Eimeria tenella project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Contact David Sibley (toxoe@borcim.wustl.edu) for further information relating to organism, libraries, or clone availability.  
Seq primer: -40RP from Gibco  
High quality sequence stop: 373.

#### FEATURES

source

1. 494  
/organism="Eimeria tenella"  
/strain="LS18"  
/db\_xref="taxon:5802"  
/clone\_lib="Eimeria tenella S5-2 cDNA Neg Selected"  
/dev\_stage="Sporozoite stage"  
/lab\_host="SOLR"  
/note="Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI ; Sporozoites were obtained from in vitro sporulated and excysted oocysts of E.tenella grown in chickens. cDNA was synthesized from poly mRNA using an oligo-dT primer containing a XhoI site. Following second strand synthesis, EcorI adapters were ligated to the cDNA and products were size-selected on sephacryl S500. The cDNA were ligated to EcorI/XhoI prepared lambda ZapII (Stratagene). Clones were converted to phagemids by mass excision using ExAssist helper phage and E.coli SOLR cell (Stratagene). Clones were selected by negative hybridization against a pool of overrepresented ESTs (N>=10, from 1682 previous reads). Insert sizes range from 1.2-2.9Kb. The library may contain a small percentage of host or bacterial contaminants."

146 a 130 c 92 g 126 t

BASE COUNT

ORIGIN

Query Match 68.8%; Score 17.2; DB 155; Length 494;  
Best Local Similarity 86.4%; Pred. No. 1.4e+03;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;



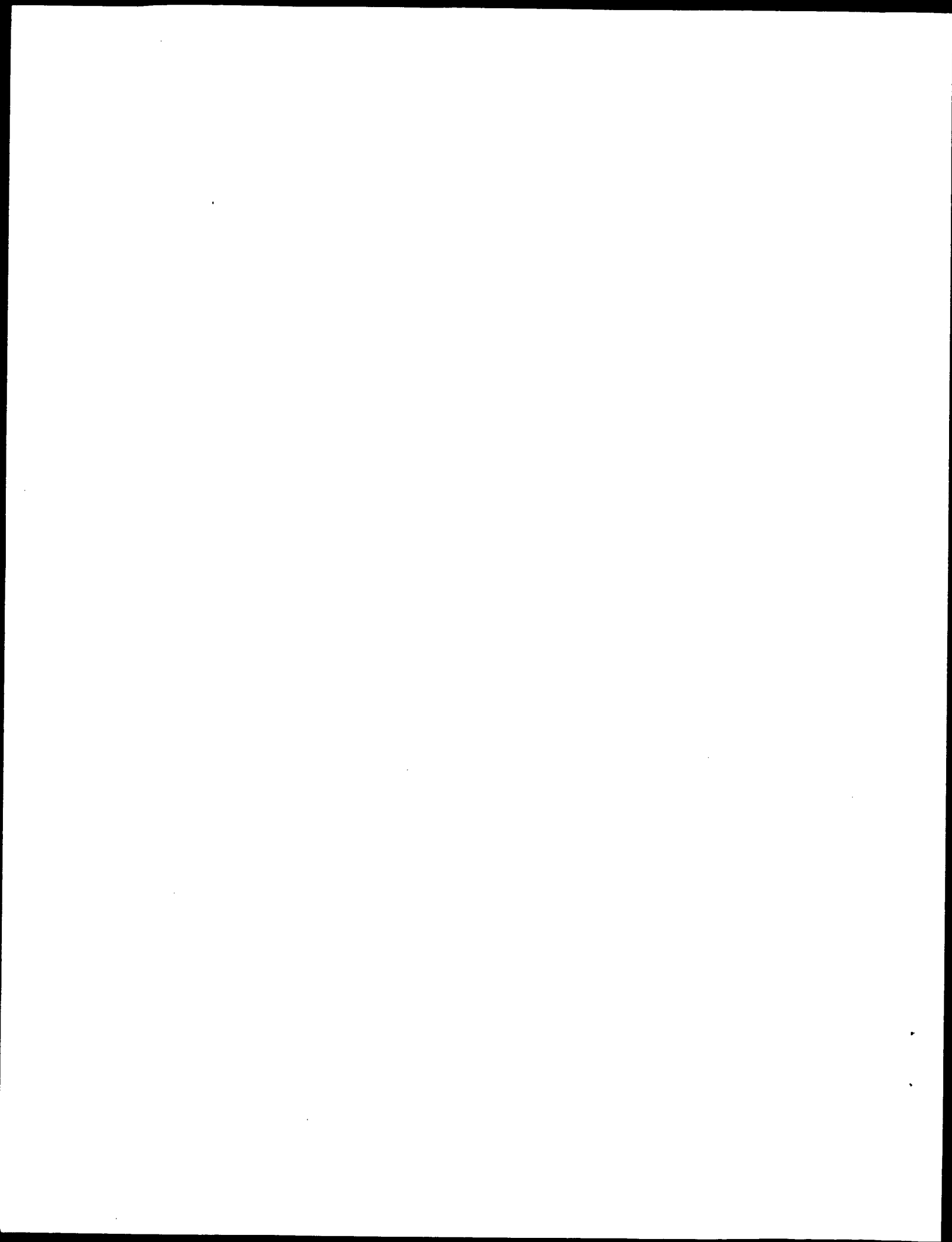
Wed Oct 10 07:44:23 2001

us-09-396-196f-10.std.rst

Page 23

Oy 4 tcaggtcaggtcagcagttg 25  
||| ||||| ||||| |||||  
Db 151 TCATGTGCAGGTCTCATCATG 130

Search completed: October 9, 2001, 13:46:55  
Job time: 9530 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 12:08:32 ; Search time 1666.31 Seconds

(without alignments)  
232.066 Million cell updates/sec

Title: US-09-396-196f-2

Perfect score: 25

Sequence: 1 attgtcgcaagtcacagaattattt 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_ba3.\*  
4: gb\_in1.\*  
5: gb\_in2.\*  
6: gb\_in3.\*  
7: gb\_on.\*  
8: gb\_ov.\*  
9: gb\_pat1.\*  
10: gb\_pat2.\*  
11: gb\_ph.\*  
12: gb\_pl1.\*  
13: gb\_pl2.\*  
14: gb\_pl3.\*  
15: gb\_pl4.\*  
16: em\_ba1.\*  
17: em\_ba2.\*  
18: em\_fun.\*  
19: em\_htgo\_hum.\*  
20: em\_htgo\_inv.\*  
21: em\_htgo\_rod.\*  
22: em\_htg\_hum1.\*  
23: em\_htg\_hum2.\*  
24: em\_htg\_hum3.\*  
25: em\_htg\_hum4.\*  
26: em\_htg\_hum5.\*  
27: em\_htg\_hum6.\*  
28: em\_htg\_hum7.\*  
29: em\_htg\_hum8.\*  
30: em\_htg\_inv1.\*  
31: em\_htg\_inv2.\*  
32: em\_htg\_other.\*  
33: em\_htg\_rod.\*  
34: em\_hum1.\*  
35: em\_hum2.\*  
36: em\_hum3.\*  
37: em\_hum4.\*  
38: em\_hum5.\*  
39: em\_hum6.\*  
40: em\_hum7.\*  
41: em\_in.\*  
42: em\_on.\*  
43: em\_or.\*

44: em\_ov.\*  
45: em\_pat.\*  
46: em\_ph.\*  
47: em\_pl.\*  
48: em\_ro.\*  
49: em\_sts.\*  
50: em\_sy.\*  
51: em\_un.\*  
52: em\_v1.\*  
53: gb\_sts1.\*  
54: gb\_sts2.\*  
55: gb\_sts3.\*  
56: gb\_sy.\*  
57: gb\_un.\*  
58: gb\_v1.\*  
59: gb\_v12.\*  
60: gb\_htg1.\*  
61: gb\_htg2.\*  
62: gb\_htg3.\*  
63: gb\_htg4.\*  
64: gb\_htg5.\*  
65: gb\_htg6.\*  
66: gb\_htg7.\*  
67: gb\_htg8.\*  
68: gb\_htg9.\*  
69: gb\_htg10.\*  
70: gb\_htg11.\*  
71: gb\_htg12.\*  
72: gb\_htg13.\*  
73: gb\_htg14.\*  
74: gb\_htg15.\*  
75: gb\_htg16.\*  
76: gb\_htg17.\*  
77: gb\_htg18.\*  
78: gb\_htg19.\*  
79: gb\_htg20.\*  
80: gb\_htg21.\*  
81: gb\_htg22.\*  
82: gb\_htg23.\*  
83: gb\_htg24.\*  
84: gb\_htg25.\*  
85: gb\_pr1.\*  
86: gb\_pr2.\*  
87: gb\_pr3.\*  
88: gb\_pr4.\*  
89: gb\_pr5.\*  
90: gb\_pr6.\*  
91: gb\_pr7.\*  
92: gb\_pr8.\*  
93: gb\_pr9.\*  
94: gb\_rol.\*  
95: gb\_rol2.\*  
96: gb\_in4.\*  
97: gb\_pr10.\*  
98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	25	100.0	1041	9	AR029499 Sequence
2	25	100.0	1041	9	AR034916 Sequence
3	25	100.0	1084	9	A11530 BioB gene o
4	25	100.0	1121	10	E00893 Genomic DNA
5	25	100.0	5793	2	J04423 E.coli 7,8-
6	25	100.0	5872	9	A38246 Sequence 1
7	25	100.0	5872	9	A38251 Sequence 6
8	25	100.0	5872	9	A93674 Sequence 1





```

/codon_start=1
/transl_table=11
/protein_id="AAA23517.1"
/db_xref="GI:145427"
/transl_table="NATYKQIAAAGRAAHYEQHADLQROSDALLAMLPOKRYT
HYLDAGCGPGWMSRWRERHAQVTDLSPPMLVOAROKDAADHYLAGDIESLPLATA
TFDLAWSNLAVQWCGNLSTALRELRYVRPKGVAFVTLVQGSLSLPERHQAQVDERP
HANRFLPDEIDQSLNGVHYQHIIQIPIITLWFDLALSAMRSLKIGICATHLHGRDPRIL
TRSOQLRQLAWPOQGGRYPLTYHLFLGVARE"
4938..5597
/gene="biob"
/gene="BIOD"
/codon_start=1
/transl_table=11
/protein_id="AAA23518.1"
/db_xref="GI:145428"
/transl_table="MSKRYEVTGDTVEGKTVASCALLOAKRACRYTAGYKPVASGS
EKTPEGLNSDALORNSQLDYATVNPYTFAPESPHIISAQGRPIESLVSAG
LRLAQADWLVEGAGWETPLSDTFTFADWVTOEOLPVILVVGKLGICINHAMLTA
QVIQHAGLTLAGWANDVTPEPKRHAETMTLTRMIPRCWERSPGLQIKQMRQPS
T"
BASE COUNT      1363 a 1554 c 1631 g 1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match      100.0%; Score 25; DB 2; Length 5793;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25
|||||
Db 2035 ATTGTCGCAAGTCACAGAATTATT 2059

RESULT 6
LOCUS      A38246      5872 bp      DNA
DEFINITION      Sequence 1 from Patent WO9408023.
ACCESSION      A38246
VERSION      A38246.1      GI:2294844
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM      Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 5872)
Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
Patent: WO 9408023-A 1 14-APR-1994;
LONZA AG (CH)
Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 950809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A-15/9"
1..96
/function="PROMOTER PTAC"
/evidence=experimental
23..28
-35_signal
-10_signal
/standard_name="PROMOTER PTAC"
/evidence=experimental

```

```

105..119
/standard_name="BIOB RBS NO.9"
/evidence=experimental
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"
/codon_start=1
/transl_table=11
/number=1
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAA02325.1"
/db_xref="GI:2294845"
/transl_table="MAHRPRWTLISOVTELFEPKPLDLLFEAQOVHROHFDPROVOVST
LLSIKTAGCPEDCKYCPQSRKYKGLAEARLMEVEQVLESARKAKAAGSTRFCMGAAW
KNPHERDMYLEOMVGVKAMGLEACMTLCTLSAQARLANAGLDIYNHNLDPSEF
YGNLITTRYOERLDTLEKVRDAGIKVSGGIVGETYKDRAGLLQLANLPTPES
VPINMLVKVGTPLADNDVDAPDFIRTAVARIMMPTSYVRLSAGREOMNEQTQAMC
FMAGANSIFYGCKLLTPNPEEDKDLQLEFKLGNLPQQTAVLAGDNEQOORLEQALMT
PDTDEYNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/number=3
/product="PROTEIN"
/protein_id="CAA02326.1"
/db_xref="GI:2294846"
/transl_table="MATVKNQKIAAAGRAAHYEQHADLQROSDALLAMLPOKRYT
HYLDAGCGPGWMSRWRERHAQVTDLSPPMLVOAROKDAADHYLAGDIESLPLATA
TFDLAWSNLAVQWCGNLSTALRELRYVRPKGVAFVTLVQGSLSLPERHQAQVDERP
HANRFLPDEIDQSLNGVHYQHIIQIPIITLWFDLALSAMRSLKIGICATHLHGRDPRIL
TRSOQLRQLAWPOQGGRYPLTYHLFLGVARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:"
8-AMINO-7-OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/number=5
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAA02327.1"
/db_xref="GI:2294847"
/transl_table="MTDDBLAFDORHWHVPTSMTPSLPVYVVSAGCELILSDGRR
LYDGMSWAAIHGYNHLPQNAAMKSQIDAMSHVFGGITHAPAEICRKLVAWTPQP
LECVLADSGSVAVEMAKMALQYQWAKGEARQELTFNGTHGDTFGAMSVCPDPS
MHSMLKGYLPENLFAPAPQSGMDGEMDERMDVGFARLMAHREHETAAVILIEPIVOGAG
GNRHYHPWLKRIKICDREGILLIADIEATFGTGRKLFACAEHAEIAPILICLGKAL
TGGTWTLSATLTITREVAETISNGEAGCEMHGPTFMGNPLACAAAASLAI LESGDWQ
QVADIEVOLREOLAPARDAEMVADVRLVGAIGVVETTHPVNMAALQKFFVEQGVWIRP
EGKLLYLMPPYITLPPQQLRLTAANRAVQDETFCQ"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/number=6

```

```

/evidence=experimental
/product="PROTEIN"
/protein_id="CAA02328.1"
/db_xref="GI:2294848"
/translation="MKLISNLDLSDGDKLPHRHVFMGMGYDGDNISPHLAWDDVPAGTK
SFVTCYDPAIDPAGTSGMWHVYVNLPAIDRVLPQFGSGIIVAMPDGYLQTRDFGKTG
YDGAAPRKGETHRYIFTVHALDIERIDVDEGASGMVGFNVHPSLASASITAMFS"
stem_loop
terminator
5583..5605
5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25
|||||
Db 140 ATTGCGCAAGTCACAGAATTATT 164

RESULT 7
A38251 5872 bp DNA PAT 05-MAR-1997
LOCUS
DEFINITION Sequence 6 from Patent WO9408023.
ACCESSION A38251
VERSION A38251.1 GI:2294849
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 5872)
Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
BIOLOGICAL METHOD OF PRODUCING BIOTIN
Patent: WO 9408023-A 6 14-APR-1994;
LONZA AG (CH)
Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 9500809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BioF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAA02329.1"
/db_xref="GI:2294850"
/translation="MSWDEKINAAIDARRADALRRRYPVAQAGRWLVADDDROYLNF
SSNDYLSHHPQIRIAWQGAQEGIGSGSGHVSIGSVYVHQALEEAEWLGYSRA
LIFISGFANQAVIAAMAKEDRIAADRLSHASLEASISPSOLRRFAHNDVTHLAR
LLASPCQQMVTVEGVFMDGDSAPLAEIQVTTQHNGLMWLDVDAHGTVIGBQGRG

```

```

SCWLOKVPBELLVTFKGVSAAVCSSTVADYLLQFARHLIVSTSMPPAQAL
RASIARVRSDEGDAREKLAALITRERAGVQDLPTLADSCSAIOPLIVGDSRALQL
AEXLROGCGWVTAIRPTTPACTARLRILTRAHEMQDIDRLLEVLHGNG"
3030..3045
/standard_name="BIOB RBS"
3043..3753
/gene="BIOB"
3043..3753
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
EKTPEGLRNSDALORNSLQLDYATVNPYTFAPTSPIILSAOGRPIESLIVMSAG
LRALEQOADMVLVEGAGGFTPLSDTFTFADWVTOQLPVILVGVKLGICINHAMLTA
QVTOHAGLTLAGVANDVTPPGKRAHEYMTTLTLMIPAPLGEIPWLAENPENATGK
YINLAFVDASTLGTSL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25
|||||
Db 140 ATTGCGCAAGTCACAGAATTATT 164

RESULT 8
A93674 5872 bp DNA PAT 22-JAN-2000
LOCUS
DEFINITION Sequence 1 from Patent EP0798384.
ACCESSION A93674
VERSION A93674.1 GI:6741862
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 5872)
Birch, O. and Brass, J.
BIOLOGICAL METHOD OF PRODUCING BIOTIN
Patent: EP 0798384-A 1 01-OCT-1997;
LONZA AG (CH)
Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A-15/9"
1..96
/function="PROMOTER PTAC"
/evidence=experimental
23..28
/standard_name="PROMOTER PTAC"
45..49
/evidence=experimental
105..119
/standard_name="BIOB RBS NO.9"
/evidence=experimental
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"
/codon_start=1

```

```

/transl_table=11
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRPRWTLQSQTVEFKPLDLLLFEAQOQVHRQHFDPQVQVST
LLSTKACPEDCKYCQSSRYKTGLEAERLMEVEQVLESARKAKAGSTRFCMGAAW
KNRHPDMPLYLEOMQCVKAMGLACMTGLTLESQAORLANAGLDYNNINLDTSPF
YGNITRTYQERLDTLEKVDAGIKVCSGIVCLGETVKDRAGLLLOLANLPTPES
VPINLVKYGTPPLADNDVDAEDFIITIAVARIMPTSYVRLSAGREQNEQTQAMC
FMAGANSIFYGCKLLTTPNPEEDKDLQLFKRLGLNFOOTAVLAGDNEQQRLEQALMT
PDTDEYYNAAAL"
RBS
2284..2297
/standard_name="BIOC RBS"
gene
2295..3050
/standard_name="BIOC"
/standard_name="BIOC"
CDS
2295..3050
/standard_name="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVKNQAIATAAFGRAAHYEQHADLQRODADALLAMLPQRKYT
HVLDAAGCPGMSRHRERHAQVTALDSPMLVQARQKDAADHYLAGDIESLPLATA
FDLWNLAVQWCGNLSLALRELYRVVRKGVVFTTLVQSGSLPELHQAQVADRP
HANRFLPDETEQSLGVHYOHHTIOPITLWFDLALSAMRSKLGIGATHLHEGRDPRIL
TRSQRLQLQLANPQOQGRPLYTHLGLVIARE"
RBS
3742..3752
/standard_name="BIOA RBS"
gene
3750..5039
/standard_name="BIOA"
/standard_name="BIOA"
CDS
3750..5039
/standard_name="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTDDLAFDORHIWHTYTSMTSPLPYVYVVSAGGCELILSDGRR
LVDMSSWAAIHCYNHNPQLNAAMKQSIDAMSHVMEFGITTHAPAIELCKRLVMTPOP
LECFLADSGSVAVVAMKALQYQAKGEARQRLTFRNGYHGDTFGAMSVCDPDS
MHLWKCYLPENLFAPOQSRMDGDERDMVGFARLMAAHRHEIAAVIIPIVQAG
GMRMYHPEWLKRIKICDRGILLIADLAIATGFTGCKLPACBHAELADILCLGKAL
TGWTWLSALTITREVAETISNGEAGCFMGFTFMGNPLACAAANASLILSDGWOQ
QVADIEVLREQLEAPARDAEMVADRVLGAIGVETHTPHVMAALQFFVEQGVWIRP
FGKLIYLMPPYIILPQOLRLTAANRAVQDETFCCQ"
RBS
5088..5100
/standard_name="ORFI RBS"
gene
5098..5574
/standard_name="ORFI"
/standard_name="ORFI"
CDS
5098..5574
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDGLKPHRRVFNMGYDGNISPHLAWDDVPAGTK
SFVYTCYDPADPTGSGMHWVNVNLPADTVLPQGFSGVLVAMPDGVLOTRDFGKTG
YDGAAPPKGTHRYITVTHALDIERIDVDEGASGAMGVNHHVHSLASIIAMES"
5583..5605
5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
stem_loop
terminator
BASE COUNT 1318 a 1552 c 1695 g 1307 t

```

## ORIGIN

```

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgtcgaagtccacagaattattt 25
|||||
Db 140 ATTGTCGAAGTCACAGAATTATT 164
|||||

RESULT 9
LOCUS A93679 5872 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 6 from Patent EP0798384.
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS Escherichia coli.
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O. and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
source
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQEKINAAIDAPRAADALRRYPVAGGAGRWLVADROYLNF
SSNDYLGSHRPOIIRAWQGAEGFGIGSGSGHVSYGYSVVHQALEELAEWLGYSRA
LLFTSGFAANQAVTAAMWAKEDRTAADRLSHASLLEAASLSPQLRFAHNDVTHLAR
LLASPCQOQVVTGEGVSMGDSAPLAEIOQVTHQNGWLMVDHGTGVYGGGRG
SCWLOKVPKELIVVTFGKGVGSGAAVCSSTVADYLIQFAHLLIYSTMPPAQAL
RASLAVIRSDGEDARREKLAALITFRFAGVODLPFTLADSCSATOPLIVGNSRALQL
AEKLRQCGWVTAIRPVPVAGTARLRLTLTAHEMODIDRLLEVLHGNG"
3030..3045
/standard_name="BIOF RBS"
3043..3753
/gene="BIOF"
3043..3753
/gene="BIOF"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYFVTGTDTEVGKTVASCALLOAKAACYTAGKYPVASSG
EKTPGLRNSDALQORNSLQDVTATVNPYTFAPESPHIISAGRGPIESLWMSAG
LRALQEQADWLVVEGAGGWFPLSDTFTFDVMTQEQPLPVLVVGWGLGCINHAMLTA
QVIQHAGLTLAGVANDVTPPKRHAETMTTLTRMIPAPLLGLGEIPWLAENPENATGK

```



```

BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN
YINIAFVDASTLGTSLR"

Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtcacagaattattt 25
    |||||
Db 140 ATTGTCGCAAGTCACAGAATTATT 164

RESULT 10
LOCUS      AR101809      5872 bp      DNA      PAT      14-FEB-2001
DEFINITION Sequence 1 from patent US 6083712.
ACCESSION AR101809
VERSION   AR101809.1 GI:12812607
KEYWORDS
SOURCE    Unknown.
ORGANISM  Unclassified.
REFERENCE 1 (bases 1 to 5872)
AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE     Biotechnological method of producing biotin
JOURNAL   Patent: US 6083712-A 1 04-JUL-2000;
FEATURES  Location/Qualifiers
            source
            1..5872
            /organism="unknown"

BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtcacagaattattt 25
    |||||
Db 140 ATTGTCGCAAGTCACAGAATTATT 164

RESULT 11
LOCUS      AR101810      5872 bp      DNA      PAT      14-FEB-2001
DEFINITION Sequence 6 from patent US 6083712.
ACCESSION AR101810
VERSION   AR101810.1 GI:12812608
KEYWORDS
SOURCE    Unknown.
ORGANISM  Unclassified.
REFERENCE 1 (bases 1 to 5872)
AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE     Biotechnological method of producing biotin
JOURNAL   Patent: US 6083712-A 6 04-JUL-2000;
FEATURES  Location/Qualifiers
            source
            1..5872
            /organism="unknown"

BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtcacagaattattt 25
    |||||
Db 140 ATTGTCGCAAGTCACAGAATTATT 164

RESULT 12
LOCUS      AE000180      11022 bp      DNA      BCT      01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete genome.
ACCESSION AE000180 U00096
VERSION   AE000180.1 GI:1786988
KEYWORDS
SOURCE    Escherichia coli K12.
ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE 1 (bases 1 to 11022)
AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,M.T., Burland,V.,
            Riley,M., Colliado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
            The complete genome sequence of Escherichia coli K-12
            Science 277 (5331), 1453-1474 (1997)
            97426617
            9278503
            2 (bases 1 to 11022)
            Blattner,F.R.
            Direct Submission
            Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
            3 (bases 1 to 11022)
            Blattner,F.R.
            Direct Submission
            Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
            4 (bases 1 to 11022)
            Plunkett,G. III.
            Direct Submission
            Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
            This sequence was determined by the E. coli genome project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@ambr.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
            Location/Qualifiers
            1..11022
            /organism="Escherichia coli K12"
            /strain="K12"
            /sub_strain="MG1655"

```

```

repeat_region
/db_xref="taxon:83333"
<i>.130
/note="REP (repetitive extragenic palindromic) element;
contains 4 REP sequences"
complement(147..1430)
/gene="ybhC"
/note="ybhC"
complement(147..1430)
/gene="ybhC"
/function="orf; Not classified"
/note="f427; 98 pct identical to fragment YBH_C_ECOLI
SW:P46130 (300 aa) but contains 127 additional C-terminal
residues"
/codon_start=1
/transl_table=11
/product="putative pectinesterase"
/protein_id="AAC73859.1"
/db_xref="GI:1786989"
/translation="MNTFSVSRLLALAFGLTLFACSSPPDPQSPDQATGTSRPI
LSAKAQNFDAHYFASLTGGAANPSPITLPAQDFVVGPAQTGVTHTTIAAYD
AAIIKTRKQYIYMPGEYQCTVYVPAAPGSIILYGTGKPIDVKIGLSLDGMSPA
DWRHVNPRGRMYKPAWYMYDSCSKRSIGLYGSAFWSQNGLQNLNIENT
LGSVDAGNHFAVALRTGDQVOIINNVLGRQNTFFVTNSGVONRLTNRQPTLVT
NSYIRGDDIVSGRAGVDFDTEPRVNSRTQOEAYFAPATLSNIYGGFLAYNSRN
AFGCGAOLGRSLVDANTNGQVIRDSAINEGENYAKPMADAVISRPFAGNTGSVD
DNDETQRNLNTNRMWMEYNNRNGVSGKVAEAKK"
complement(1507..1535)
/note="factor Sigma70; predicted +1 start at 806574"
complement(1582..2058)
/gene="ybhB"
/note="ybhB"
complement(1582..2058)
/gene="ybhB"
/function="orf; Unknown"
/note="f158; 99 pct identical to YBH_B_ECOLI SW: P12994"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC73860.1"
/db_xref="GI:1786990"
/translation="MKLISDLRDGDKLPHRHVFMNGYDGNITSPHLADWDVPAGTK
SFVTCYDPDPTGSGMHWVYVNLPAIDRVLPQFGSLVAMPDGVLPQRTDFGKTG
YDGAAPRGKTHRYFTVHALDIERIDVDEGASGMVGFVHFSLASAITAMFS"
complement(2108..2124)
/note="central position to predicted promoter:85.5"
/bound_moiety="RnaS predicted site"
complement(2117..3406)
/gene="bioA"
/note="b0774"
complement(2117..3406)
/gene="bioA"
/EC_number="2.6.1.62"
/function="enzyme; Biosynthesis of cofactors, carriers:
Biotin"
/note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
/codon_start=1
/transl_table=11
/product="7,8-diaminopelargonic acid synthetase"
/protein_id="AAC73861.1"
/db_xref="GI:1786991"
/translation="MTTDDLAFDQQRHITWHPYTSMTSPLPYVPVVAEGCELLISDGR
LVDMSSWMAAIIHGYNPOLNAOKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
LECFLADSGSAVEAVAKMALQYQAKGARQRFILTRNGYHGDTFGAMVCGDPDS
MHLKGLYPENLFAPAQSMQMDGEMDERMVGFARLMAAHRHEIAAIIETPVOCAG
GMRMYHPKWLKRIKICDRGILLIADEIATGFRGCKLPACBHAETADLILGKAL
TGGTMTLSATLTREVAETISNGEACFPHGPTFMGNPLACAAANASLAILESGDWQQ
QVADIEVLREQLAPARDAEMADVRLVIGAGVGVETTHPVHMAALQKFFVEQGIWRP
FGKLIILMPYIILPQQLRLTAANRAVODETFCCQ"
complement(2193..2221)
/gene="bioA"
/note="factor Sigma70; predicted +1 start at 807260"
3411..3450
/note="central position to bioB promoter: -20"

protein_bind
/bound_moiety="BioB documented site"
3411..3450
/note="central position to predicted promoter: -20; genetic
evidence for the site"
/bound_moiety="BioB predicted site"
complement(3411..3450)
/note="central position to predicted promoter:50"
complement(3411..3450)
/note="central position to bioA promoter:50"
/bound_moiety="BioB documented site"
3413..3441
/note="factor Sigma70; promoter bioB; documented +1 at
808525"
complement(3447..3473)
/note="factor Sigma70; promoter bioA; documented +1 at
808515"
3493..4533
/gene="bioB"
/note="b0775"
3493..4533
/gene="bioB"
/EC_number="2.8.1.1"
/function="enzyme; Biosynthesis of cofactors, carriers:
Biotin"
/note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
/codon_start=1
/transl_table=11
/product="biotin synthesis, sulfur insertion?"
/protein_id="AAC73862.1"
/db_xref="GI:1786992"
/translation="MAHRPRWTLQSQVTELFERPLLDLLEAQQVHQHFDPQVOVST
LISIKTGACPEDCKYQSSRYKTLGLEAPRLMEQVLESARKAKAAGSTRFCMAAAM
KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSQAOPLANAGLDYVNHNLDTSEP
YGNLITRYQORLDTLEKVRDAGIKVCSGGIVGIVKDRAGLLQLLANLPTPEP
VPINMLYKVGKPLADNDVDADFIRTIATVARIIMPTSYVRLSAGREOMNEQOAMC
FMAGANSIFYGCKRLTTPNPEDKDLQFLRGLNLPQQTAVLAGDNEQQORLEQALMT
PDTDEYNAAL"
4530..5684
/gene="bioF"
/note="b0776"
4530..5684
/gene="bioF"
/EC_number="2.3.1.47"
/function="enzyme; Biosynthesis of cofactors, carriers:
Biotin"
/note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
/codon_start=1
/transl_table=11
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAC73863.1"
/db_xref="GI:1786993"
/translation="MSWQEKINAAILDARRAADALRRYPVQAQAGRWLVADDDRYLNF
SSNDYLGISHHPQIIIRAWQOAGAEQFGGGGSHVSVVYVQALFEELAEMLGYSRA
LLFISGFAANQAVIAAMAKEDRIAADRLSHASLEAASLSPSOLRFAHNDVTLAR
LLASPCGOQMVVTEGVSMGDSDAPLAEIQVQTOOHNGWLMVDAGHTGVIGQGRG
SWLQKVPPELLVITFGKFGVSGAAVLCSSVADYLLQFARHLIYSTMPMPAQAL

Query Match 100.0%; Score 25; DB 1; Length 11022;
Best Local Similarity 100.0%; Pred. No. 0.088;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgtcgcaagtacagaattattt 25
|||||
Db 3516 attgtcgcaagtacagaattattt 3540

RESULT 13
AE005258
LOCUS AE005258 13501 bp DNA BCT 21-MAR-2001
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
of 155.
ACCESSION AE005258 AE005174

```

```
VERSION      AF005258.1  GI:12513751
KEYWORDS
SOURCE       Escherichia coli O157:H7 EDL933.
ORGANISM     Escherichia coli O157:H7 EDL933
              Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
              Escherichia.
REFERENCE    1 (bases 1 to 13501)
AUTHORS      Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
              Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
              Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
              Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
              Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
              Welch,R.A. and Blattner,F.R.
TITLE        Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
JOURNAL      Nature 409 (6819), 529-533 (2001)
MEDLINE      21074935
PUBMED       11206551
REFERENCE    2 (bases 1 to 13501)
AUTHORS      Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
              Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
              Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
              Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
              Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
              Welch,R.A. and Blattner,F.R.
TITLE        Direct Submission
JOURNAL      Submitted (22-OCT-2000) Laboratory of Genetics, University of
              Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES     Location/Qualifiers
              1..13501
               /organism="Escherichia coli O157:H7 EDL933"
               /strain="EDL933"
               /serotype="O157:H7"
               /db_xref="taxon:155864"
               /note="enterohaemorrhagic"
               <1..7576
               /note="O-island #36; Region of the EDL933 chromosome not
               homologous to E. coli K-12 MG1655; Cryptic prophage
               CP-933k; includes one copy of the 13 bp direct repeat that
               flanks the prophage"
               66..665
               /gene="lomk"
               /note="lomk"
               66..665
               /gene="lomk"
               /function="putative membrane; Other or unknown (Phage or
               Prophage Related)"
               /note="Residues 1 to 199 of 199 are 72.86 pct identical to
               residues 1 to 199 of 199 from GenPept 118 :
               gi17532789|gb|AAF63231.1|AF151091_2 (AF151091) Lom
               [prophage P-EibA]"
               /codon_start=1
               /transl_table=11
               /product="putative outer membrane protein of prophage
               CP-933k"
               /protein_id="AAG55137.1"
               /db_xref="GI:12513752"
               /translation="MRKVCAILSAATCLAVSGVPANASHQSTLSAGYLHASTDAPG
               SDDLNGINVKRYETDTLGLITSFYNAEDEQKTHYSOTRWHEDEVRRWFVSMAG
               PSVRVNEWFSAYAMAGVAYSRSVTSFGDYFTRKTKTHDVLTSDDARSNTSLAW
               GAGQFNPTESVADVAVAYEXSGSGDWRTDGFIVGQYKF"
               719..2041
               /gene="Z0982"
               719..2041
               /gene="Z0982"
               /function="putative structure; Structural component (Phage
               or Prophage Related)"
               /note="Residues 164 to 440 of 440 are 68.79 pct identical
               to residues 381 to 645 of 645 from GenPept 118 :
               gi14585437|gb|AAD25464.1|AF125520_59 (AF125520) putative
               tail fiber protein [Bacteriophage 933W]"
               /codon_start=1
               /transl_table=11
               /product="putative tail component of prophage CP-933k"

misc_feature
source
1..3469
/gene="Z0985"
2489..3469
/gene="Z0985"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 13 to 326 of 326 are 55.73 pct identical
to residues 18 to 331 of 336 from GenPept 118 :
gi16960367|gb|AAF33527.1| (AF170176) hypothetical protein
predicted by Glimmer [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/translation="MLSPRTTTFHNSVNIQSSPQTVSFAGKEYELKVIDEKTPIFL
QWEPNPERYKKDEVPVNTKQHPYLDNVTNAARIESDRMIGIFVDGDFSNQKTAFLS
KLERDENWIIYFREDVDFSMYDRKSLDIYHDIICEQRLTEDKRDREYLLNLEKELR
ETSKAQDSLISMYAKRNHAWDFDFRNALLKAGEIFRCTYNTKNGHISFGGCIYLD
MDMLTGLKGTIYADGISHVDRNDSYNIENSALIVNRSNHPALLEGISPMHSKVD
AHPYDGLGKGVKKYFNFTPLHYNHFCDFIEFNHFNITMNTSQYTCSSW"
3530..4522
/gene="Z0986"
3530..4522
/gene="Z0986"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933k"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/translation="MKIPSLQSNFNSAPAGYSAPIAPNAENAVADYVLDIGKRIPL
SAADLSNYESVIRAVDRSRLLIDQHTVDMICNTVLDALSRQFADSVYGIINEK
VHIGCIKRYNEELNEESSVKIDDIQSLTCNELYEDVQCELPFICEAGENDEPEY
VSPSAPDSDSYEMPQGLIHIIHHVTGSDSPSGDSNIGELTEILLARVAQELG
WSVPDFKGYAEPEERAEHLRLNLNLRQAAMHEENERAFFELGTISDRYBASPDFT
EYSAVSNIGYGFIIQQHDFPGLAINDLNQANQILYHGAPYIFTFGVDKHNQ"
5350..6231
/gene="Z0989"
5350..6231
/gene="Z0989"
/function="orf; Other or unknown (Phage or Prophage
Related)"
```

```

/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55142.1"
/db_xref="GI:12513757"
/translation="MLSPYSVNIAGCSWNSLTNRNLTSPDNRVLSVRDAAVHSDNGAOV
KVGNTYRVATDNKFCVTRSHSGCFNLLHRLGWPKEGJSRKIEVLMNAPVSAAM
ERGIYHSRNPDPVPPDYAPPPELPSVDYVNLSPGCVGNVYEDAEADATYVKM
FTTSSQNEVSEVCEFCNQYVAGSAEKIYGNNGDIIIGIRMDKINGESLLNISLPAQ
ABHAIYDFDLRQKGLFVDTETNVLVYDRAKNEFNPIDISSNVSDRSWSESQIMQ
SHGKQDLISVVLSKI"
/length=6462
/genes="20990"
/length=6462
/genes="20990"
/function="orf; Other or unknown (phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55143.1"
/db_xref="GI:12513758"
/translation="MRPISLNLVHQSSRSSMSDSDIESLVKASSVQWTKNNPQLRF
QCTDHNITQQIEAALDKRGSTGTGRVLLNAIESISRLKSETVWHLNLSRIGVMAHRD
IDAENHRTGSDHCNNAVEYPCGEGISVYDFHATVFFHELHVNHLNGERLKVES
SRAESQKSPILLBEARTVGLGAFSEEVLSNKFHEIGMPRTSYDPSXSAIHDDNT"
VSLGFOQVRLHPL"
/complement(7668..8144)
/genes="ybhB"
/length=8144
/genes="ybhB"
/complement(7668..8144)
/function="orf; Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="MKLISNLDLDRGDKLPHRVFNGMGYDGDNISPLHAWDDVPACTK
SFVTCYDPAPTGSGWVNVNLPADTVLPGCGSLGVPDGLVQRTDFGKIG
YDGAAPKGEHRIFTVHALDVERIDVDEGASGVVGNVHFHSLASATIMAFS"
/complement(8203..9492)
/genes="bioA"
/note="Z0993"
/complement(8203..9492)

Query Match 100.08; Score 25; DB 1; Length 13501;
Best Local Similarity 100.08; Pred. No. 0.088;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgtcgcaagtccagcaattatttt 25
|||||
Db 9602 ATTGTGCGAAGTCACAGAAATTATT 9626

RESULT 14
AP002553
LOCUS AP002553 297816 bp DNA BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION AP002553 BA000007
VERSION AP002553.1 GI:13360211
KEYWORDS
SOURCE
Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (sites)

```

## AUTHORS

Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S., Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T., Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and Shinagawa, H.  
Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak  
Genes Genet. Syst. 74 (5), 227-239 (1999)

## TITLE

## JOURNAL

## MEDLINE

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

1..297816 Location/Qualifiers  
/organism="Escherichia coli O157:H7"  
/strain="O157:H7"  
/sub\_strain="RIMD 0509952"  
/db\_xref="taxon:83334"  
79..1245  
/gene="ECs0753"  
79..1245  
/gene="ECs0753"  
/note="similar to SUCC\_ECOLI g11786948 percent identity 100 in 388 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/transl\_table=11  
/evidence=not\_experimental  
/product="succinyl-CoA synthetase beta subunit"  
/protein\_id="BAB34176.1"  
/db\_xref="GI:13360212"  
/translation="MNLHEYQAKQLFARYGLPAPVGYACTTPREAEAEASKIGAPWV  
VKQVHAGRGKAGGVKVVNSKEDIRAEFNWGLKRVLYTOTDANGQVNOILVFAAT  
DIARELYGVAVDRSSRRVFMATSTEGVEIEKVAETPHLHKVALDPLTPMPYQG  
RELAPKLEGLKLVQQFTKIFNGLATIFLDRDLALIEINPLVITKQGLDCLDKLGA  
DGNALFROPDLREMHDQSDQEPREAAQAWELNVYVLDNGVNGVAGLAWGTMIDIV  
KLHGGEPAFLDVGGGATKERVTEAFKIIISDDKVKAVLVNIFGIVGSLADIADGIIG  
AAVEYGVNVPPVVRLEGNALGAKKGLADSLINIIAAGLTDAAQQVVAAVEGK"

```

1245..2114
/gene="ECs0754"
CDS
1245..2114
/gene="ECs0754"
/notes="similar to SUCD_ECOLI gi11786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/transl_table=11
/translation="MSLIDKNTKVICQGTGSGQTFHSEQAIAIYATKMGVTPGKG
GTHLGLPVNTREAVATGATASIVYVAPPCDKSILEAIDAGIKLIITTEGIPT
LDMITVKLDEAGVRMIGPCVITPGECKIGIOPGHHRPKGVIYRSGLTYE
AVKQTDYFGQSTCGVIGDPIPGSNFIDILEMFEQPTETAIYIMIGIISABEEA
AAVKEHVTPVVGYYAGVAPKRGMRHAGAIAGGKGTADKFAALEAAGVKTVRS
LADIGALKTVLR"
2781..3686
/gene="ECs0755"
CDS
2781..3686
/gene="ECs0755"
/notes="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcriptional activator (gcv operon activator) -
Escherichia coli gi1417043 [sp|P32064|GCVA_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/transl_table=11
/translation="MRGKIPKTELVTFFVVARHESYTRAAEELATQSAVFRQVSAL
EEELRTPLFHSKRRIPNDTGYILGYIVKETLNKLERDNTMTWQTVQVIELAVN
PTFTHLNLNHEFTKLPDIIVNIHSLANNGDFLNREYDAVIMRENFCAWAEVEY
LFREIPLVCSGLSAMDOKLSVAELLTEPLLHOSRITGTWEENFALSGVSSPLVN
NGREFDLISLIAVRSNLGVALLPRAIQLDLDSDMWIPCDVPRTGNRTMTWQE
EKSDSPHLQOFREWLAKSVVPOEM"
complement(3720..4322)
/gene="ECs0756"
CDS
complement(3720..4322)
/gene="ECs0756"
/notes="probable cob(I)alamin adenosyltransferase, similar
to cob(I)alamin adenosyltransferases (corrinoid
adenosyltransferases) e.g. [Escherichia coli]
gi1151481 [sp|P13040|BTUR_ECOLI percent identity 67 in 200
aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="putative cob(I)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARISTHVRVQKQLKEQVDRVAAAEKKGILIVFTGNGKG
KSTAAFTVTRAVGCHGTGVGAOYIKGQWNGEYNLLQPLGVFEHMGTFWTETONR
QADIDAKEVSESKRMADKRYDLVDELTYMLAVHYLDTEVVLASLQNRRAQGSV
IVTGRGCHSLIKMADTVSEIRPVKHFADNGIQAPGIDW"
complement(4332..5984)
/gene="ECs0757"
CDS
complement(4332..5984)
/gene="ECs0757"
/notes="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarase) - Escherichia coli
gi120598 [sp|P00923|FUMA_ECOLI percent identity 68 in 545
aa"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"

```

```

gene
CDS
gene
CDS
gene
CDS

```

```

/translation="MSKPFIIWOEPFLQKGTETYLISDQHTITVELDGEVIAKIAPE
ALTILARQAFVASEFLLRSALHQVVASILNDQOASSNDKVALLOLRNAEVSAGKVLQ
NCODTGTATIVASKGQOQWGTGDAEALSKGIYSTFOENLRFSONAPLDMYETVNTQ
TNLPACIDISAVAGDEYHFLCVNKGSGSANKAALYQETKSLLOPEKLTAFLEKMKSL
EASQPGGIAQFGKGYFAHDIVIRLPRHGGSCPTAMALSCSADRNKAKTNKHGVLW
EKLEHNPQGYIPASLREENHAQHVLQDLNRLPDMQDLARLPVGTSLSGPIVVAR
DIAHAKIKARLDGSEPMPEYLKHHIVYVAGPAKTPENMACSLGPTTGGRMGDIYDF
QAAGSLVMSLKGNSQOQVTDACHKHGNGLSIGGAALLAQAEVKSRLCLEVPGLG
MEAVHMEVENLPAPLIVDDKGNFFSQFEQHQRCASCAPAGH"
complement(6092..7372)
/gene="ECs0758"
CDS
complement(6092..7372)
/gene="ECs0758"
/notes="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
gi121467 [sp|P24943|GLTT_BACT percent identity 38 in 416
aa, also similar to C4-dicarboxylate transpor"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKKISLTMTMILLALVLMIGVILNNTASPETAKLYAQEISIFT
TIFLRILKMIAPLVVSTLVGIAKMGDAKALGRIFSKTFLFICASLSLIALGLITY
NFFEMGTGINFVAHGAETTGVAAPETLVKVFISHAFPTSIDAMAHNEILQIVVFSI
FLGCSLTAIGEAGSAIVHALDSIAHMLKLGIVMLFAPLVFAISALIERGLAYM
VSAGIFMGFEYFTMLLWLLGLAIVYVPCIRKLRLALSEPALLATTSSEAAFP
LTLEKLEQFGVSPKIASFVLPIGYFNLSMAYGSFATVTAQACNHLHLSIGEOITM
LLILMLTSKMGAGVPRASVMVTAATLNQFNPEAGLILLMGVDPFLDMGRSATNVMS
ANGAMVSWEGEGEGECRGKALKPNESNALP"
complement(7533..7850)
/gene="ECs0759"
CDS
complement(7533..7850)
/gene="ECs0759"
/notes="unknown"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"

```

```

Query Match 100.0%; Score 25; DB 2; Length 297816;
Best Local Similarity 100.0%; Pred No. 0.097;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtcacagaattattt 25
|||||
Db 92334 ATTGCGCAAGTCACAGAATTATT 92358

RESULT 15
AF250776
LOCUS
DEFINITION
AF250776 5526 bp DNA BCT 31-JAN-2001
Uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
modC-bioA intergenic region, DAPA-aminotransferase BioA (bioA),
biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
biotin biosynthesis protein BioC (bioC) genes, complete cds; and
dethiobiotin synthetase BioD (bioD) gene, partial cds.
AF250776
ACCESSION
AF250776.1 GI:12620124
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
20575196

```

```

11133432
2 (bases 1 to 5526)
REFERENCE Entcheva,P., Liebl,W. and Streitt,W.R.
AUTHORS Direct Submission
TITLE Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
JOURNAL Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES
    Location/Qualifiers
        1..5526
            /organism="uncultured bacterium pCosHE2"
            /db_xref="taxon:143797"
            /clone="pCosHE2"
            /note="unknown organism, cosmid clone derived from
            environmental consortium"
            complement(52..528)
            /note="ORF1"
            /codon_start=1
            /transl_table=11
            /product="hypothetical 17.1 kDa protein in modC-bioA
            intergenic region"
            /protein_id="AAG60577.1"
            /db_xref="GI:12620125"
            /translation="MKLISNDLRDGDKLPKRVHFMGMDGNISPHLAWDDVPAGTK
            SVVTCYDPAPTGSGMWHVNVNLPADTRVLPQGGSLVAMPDGVLTQTFDFCKAG
            YDGAAPKGETHRXIFTVHALDIERIDVDEGASGMVGFVHFLSASITAMFS"
            complement(587..1876)
            /gene="bioA"
            /complement(587..1876)
            /note="7,8-diaminopelargonic acid
            synthetase-aminotransferase"
            /codon_start=1
            /transl_table=11
            /product="DAPA-aminotransferase BioA"
            /protein_id="AAG60578.1"
            /db_xref="GI:12620126"
            /translation="MTTDDLAFQDQRIHWPYTSMTPLVPVYVVSABGCELIISDGRR
            LKGVSSWAAIHGYNHNPOLNAAMKSOIDAMSHVMFGGITHAPALEICRLKLVAMTPQ
            LVGTFADSGVAVEMAKALQYQAKGEARORLFRNGYHGDTFGMSVCDPDNS
            MSLWKVLPENLAPQSRMDGEWDERKVGFAARMAHRHEIAVIEPIVOGAG
            GRMYHPWLRIKIKCDREGILLIADIETFGRTGKLFACAEHAEIAPDILGKAL
            TGGMTLSATLTTEVAETISNAGGCFMGPTFMGNPLACAAANASLAIIESGDWQ
            QVADIEVLRLQAPARDAEMADVRLVIGAVGVETTHPVNMAALQRFVEQGVWRP
            FGKLIYLMPPYIILPQOLRLTAANRAVDQTEFFC"
            1963..3003
            /gene="bioB"
            1963..3003
            /gene="bioB"
            /codon_start=1
            /transl_table=11
            /product="biotin synthase BioB"
            /protein_id="AAG60579.1"
            /db_xref="GI:12620127"
            /translation="MAHRPRTWISQVTELEKPLDLLFPAQVVRHQHFDPPQVQVST
            LLSIKGACPDCKYCPSSRYRTGLEARLMEVQVLESARAKAAGSTRFCMGAAW
            KNPNERDMPYLMQVQVKGDLGACMTLGLTLESQARLANAGLDYINHLDTSPF
            YGNITTRTQERLDLTKKVDAGIKVCVSGGIVGLGTVKDRAGLLILANLPTPEP
            VPIMLVKVKTPLADNDYDAEDFIETIAVARIMPTSYVRLSAGREQNEQTQMC
            FMAGANSIFGCKLLITPNPEEDKDLQLFRKLGNLPQQTAVLAGDNEQQORLEQALMT
            PTDDEYNAAL"
            3000..4154
            /gene="bioF"
            3000..4154
            /gene="bioF"
            /note="8-amino-7-oxononanoate synthase"
            /codon_start=1
            /transl_table=11
            /product="KAPA synthetase BioF"
            /protein_id="AAG60580.1"
            /db_xref="GI:12620128"
            /translation="MSWQKINAAIDARRAADLRRLRRYPVAGAGRWLVADDDROYLNF
            SSNDYLGSHHPQIIRAWQGAQFQIGSGSGSHVSGYSVHQAALKEELAEWLIGYSRA
            LLFTSGPAAQNAVIAAMMAKEDRIADRSLASLLSAALSPQLRRFAHNDVTHLAR
            LLASPCPGQQLLVTEGVFSMDGDSAPLAETIQVQVTOQHNGWMLVDDAHGTGVIGQGRG
            SCWLQKVPPELLVVTCKGFGVSCAAVLCSSTVADYLLQFARHLIYSTMPPAQAQAL
            RASLAVIRDEGDAREKLVSLIAHFRAGVODLPFTLADSCSAIQPLIVGDNRSALQL
            AEKLRQCGCWTAIRPPTVPAGTARLRLTLTAHEMODIDRLLEVLHGNG"
            4141..4896
            /gene="bioC"
            4141..4896
            /gene="bioC"
            /note="reaction step prior to synthesis of pimeloyl-CoA"
            /codon_start=1
            /transl_table=11
            /product="biotin biosynthesis protein bioC"
            /protein_id="AAG60581.1"
            /db_xref="GI:12620129"
            /translation="MATVVKOATAAAFGRAALYEOHADLQSDADVLAMLPKRYT
            HVLDAAGCGPQWTRHNRERHAQVTLDSLPPMLVQARQKADADHYLAGDIESLPIATA
            TELDANSNLAVQWGNLSIALRELYRVRSKGVVAFTLLVQGSLLPELHQAWQAVDERP
            HANRFLPPDEIEQSLNGVYHHIQIPITILWFDDALSAMRSLKIGATGHLHEGRDPRIL
            TRSQLRLQIAMPQQQGRYPLTYHLFLGVIARE"
            4883..5526
            /gene="bioD"
            4883..5526
            /gene="bioD"
            /note="DHB synthetase"
            /codon_start=1
            /transl_table=11
            /product="dethiobiotin synthetase BioD"
            /protein_id="AAG60582.1"
            /db_xref="GI:12620130"
            /translation="MLVSKRYFVTGDTVEGTVASCALLQAAKAAGYTAGYKPYAS
            GSEKTEGLRNSDALQNRSSQLDYATVNPYTFAPTSPHIIISAOGRPIESSVMS
            SGLRAEQQADWLVVEGAGGWFPIJSDTTFADWVTOQLPVLVGVKLGICINHAML
            TAAIQHAGLITLWGVANDVTPPKRHAETITLTMIPAPLGEIPWLAENPE"
            BASE COUNT 1274 a 1507 c 1567 g 1178 t
            ORIGIN

Query Match 96.0%; Score 24; DB 2; Length 5526;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ttgtcgcaagtcacagaattattt 25
    |||||
Db 1987 TTGTCGCAAGTCACAGAATTATTT 2010

RESULT 16
AF248314
LOCUS
DEFINITION
    AF248314 8227 bp DNA BCT 24-JAN-2001
    Uncultured bacterium pCosAS1 urocanase-like protein (hutU) gene,
    partial cds; histidine ammonia-lyase-like protein (hutH), DAPA
    aminotransferase BioA (bioA), biotin synthase BioB (bioB), 7-KAPA
    synthetase (bioF), biotin biosynthesis BioC-like protein (bioC),
    and dethiobiotin synthase BioD (bioD) genes, complete cds; ABC
    transporter-like protein (elsa) gene, partial cds; and unknown
    gene.
    AF248314
    VERSION AF248314.1 GI:12407610
    SOURCE uncultured bacterium pCosAS1.
    ORGANISM Bacteria; environmental samples.
    REFERENCE 1 (bases 1 to 8227)
    Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streitt,W.R.
    Direct Cloning from Enrichment Cultures, a Reliable Strategy for
    Isolation of Complete Operons and Genes from Microbial Consortia
    Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
    11133432
    REFERENCE 2 (bases 1 to 8227)
    Entcheva,P., Liebl,W. and Streitt,W.R.
    Direct Submission
    Submitted (21-MAR-2000) Mikrobiologie und Genetik, Universitaet
    Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
    Location/Qualifiers

```

```

source
1. .8227
/organism="uncultured bacterium pCosAsi1"
/db_xref="taxon:143793"
/clone="pCosAsi1"
/note="unknown organism, cosmid clone derived from
environmental consortium"
<1. .502
/gene="hutu"
<1. .502
/gene="hutu"
/note="putative urocanate hydratase"
/codon_start=2
/transl_table=11
/product="urocanase-like protein"
/protein_id="AAG53593.1"
/db_xref="GI:12407618"
/transl_table=11
/translation="PRERISFOGLPARICWGLEWRQKLGIAFNEMVRGEVAPIVI
GRHLDSGVASPNRETEAMQSDAVSDWPLNALLNTASGATWVSLHHGGVGMGF
SOHAGMVICDGTDDAARIARVHNDPATGVMRHADAGYDIAIDCAKEQGLNLPMAI
ATGGQH"
502. .2031
/gene="huth"
502. .2031
/gene="huth"
/note="putative"
/codon_start=1
/transl_table=11
/product="histidine ammonia-lyase-like protein"
/translation="MNAULTPTLTLAOLROVMQOPLTLTDESHEAINDSVACVE
AIVAGERTAGINTGGLLAQRIATHLDENLQSLVLSHAACVGEPLDDDDIVRLMMV
LKTNSLARGFIRSVIOALIALVNAVGYSDVPKAGSVGSDGLAPLAHMSLTLLGE
GRAYRGELPRATLQAGLAPLTAKEGALLNGTQASTAFALRGFEADLFAS
AVCVGALTTEALRRPFDAPIHEVRGQGOIDAAALFRHVLTDTSATSAESHNCDK
VDPYSLRCQPOVMGACLTOMRQVAVLVLVESNAVSDNPLVFAAENEMVFRGNFAEP
VMAADNLATAICAGLERRTALMDKHMQLPPLVFRNGSVNGEMIAOVTAAL
ASENKGICHPHTISVDKPCRPANQEDHVSMAFAGRLWEMAGTRGVLAWEMLAAC
L"
QGANLDRGUTSSPLLEQARQSCGEQVARYDDDRFFAPDIEAAISLILNKGSLVGLLPAP
complement(2130. .2606)
/note="ORF1; similar to hypothetical 17.1 kDa protein in
modc-bioA intergenic region"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG53587.1"
/db_xref="GI:12407612"
/transl_table=11
/translation="MKLISQDLRDGKPLRHVFMGMVGEQDNISPHLAWDEVPSGK
SFVYCYDDPAPDTSQWVHWIIVANIPADTVLPQSGSSILVALPEGAVETRTDECKAG
YGAAPKGETHYITFTVHALNVERIEVDEEASGAMVGNVHFTLGSASIIAMYS"
complement(2651. .3940)
/gene="bioa"
complement(2651. .3940)
/gene="bioa"
/note="7,8-diamino-pelargonic acid aminotransferase"
/codon_start=1
/transl_table=11
/product="DAPA aminotransferase BioA"
/protein_id="AAG53588.1"
/db_xref="GI:12407613"
/transl_table=11
/translation="MTTDDIAFDROHIWHPYTSMTOPLPVPVVEAHGCELTIASGEQ
LVDCVPSADSGVAVEAMKMAQIYWHAKGESRQRTIFRNGYHGDFTFGAMVCDPNS
MHSLLWKGYPENLFAPOQSFDCGDMVGLAPMAAHEHIAVILEPVOGAG
GMRYHPKRLRMCDREGILLIADETATGRTGKLFACFACHTPTDILCLGKAL
TGGTMTSATLTTTQVATETISNGEAGCFMHGPTFMGNPLACAVASESLALLEGEWOD
QVAATESQOEELAVADESEFADRVAVLGAIGVETHPVNMALQORFFVARGVWVRP
FGKLLYLPPIVIRPDLSKLTTRAVDVAQAPAHFTR"
4029. .5069
/gene="biob"
4029. .5069
/gene="biob"

gene
1512 a 2642 c 2605 g 1453 t 15 others
BASE COUNT

```

```

/codon_start=1
/transl_table=11
/product="biotin synthase BioB"
/protein_id="AAG53589.1"
/db_xref="GI:12407614"
/translation="MAHQARWTLQSQTALFEKPLLELFFAQIHRQHFDPQOIQYST
LLSTKACPEDCKYCPOSARYKTGLESERMEVEQESARQAKNAGSTRFCMGAAW
KNPHERMPYLEOMVGVKAMGLOACMTLGTLDHOAQRLASAGLDYNNINLDTFPEF
YGNLTTRTYQERLDTLDKVRDAGIKVCSGIVGLCTFTVDRAGLLQLLANLPTPES
VPIINLVKGTPLADNDDVAFDXIRTXSVAXMXTXLVRLSTGEQNEQTMQC
FMAGANSIFYGCKLLTTPNPEEDKQVOLFRRKLGNLGNPOOTAVLTGDNQQHLEQQLFN
ADTDQFYNAATV"
5066. .6223
/gene="bioc"
5066. .6223
/gene="bioc"
/note="7-keto-8-amino-pelargonic acid synthetase; similar
to BioC"
/codon_start=1
/transl_table=11
/product="7-KAPA synthetase"
/protein_id="AAG53590.1"
/db_xref="GI:12407615"
/translation="MTWQORIDAALDARRAADALRVRTPVENGAGRELIREQRYCNF
SGNDYLGLSQHPETIRAWQOAGQYGVSGSGSHVGYTHAQALEEHLGWLGYPPRA
LLFISGFAANOALIAALTDKEDRIVADRLSHASLEAASHPAOLRRFAHNDVAQLDS
LLDKPCHQQLVWTEGIFSMGDSAPLVGTHAAQROGSLVVDVDAHGIGVTPGEGRG
SAASQGVKPELLVYTFRGVGVGAAVLCSDAVDYLLQFAHLLIYSTMPAGAVAL
LAFNVRINSEDNARRORLAALIDQFRAGVRDLPVQITASQSLQPLIVGENARALHL
AQRIRERIGWTVTAIRPPTVPAGTARLRLTLTAAHQPADITALVEALHVAQE"
6207. .6962
/gene="bioc"
6207. .6962
/gene="bioc"
/codon_start=1
/transl_table=11
/product="biotin biosynthesis BioC-like protein"
/protein_id="AAG53591.1"
/db_xref="GI:12407616"
/translation="MSQVKNQALAAAFGRAGQSYSRHDALQROSDGRLALLAENTEA
SVLDAGCGPGNSLNAWROTSVVTAMDLCAPMLEEARNRQAADSYLQADIEAIPLADA
QFSLAWLAVOXCASLPRAELMRVTGLGKVAFVTLNGLSPLQNAQWAVDTOP
HANRFLSQGVNDALSGWQYRCVQVTVTLFDDALSAHMSLKGIGATHLHAGRAAKPL
TRGLOQLAAQAWPQAGKFPVLSYQLFHGIIERD"
6940. .7680
/gene="bioc"
6940. .7680
/gene="bioc"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthase BioD"
/protein_id="AAG53592.1"
/db_xref="GI:12407617"
/translation="MELSNVTERYVTCGTDTVEGKTVASTALLQARALGRHTAGYKP
VAGSEMTAQGLRNSDALQNRNSLALSFAVNPYTFAPETSPHVSADQRPPIEA
TLSAGLRALETQADMVLVEGAGWFTPLSQEQTFADWYMAERPLVILVGVKLCGINH
AMLTAAQAIROAGLIFAGWGVVPGKRAEYMATLRRVLSAPLGLGIPWLGQADT
SAVGQYLDLTALTSPAPSOGOSAGHPAR"
complement(7619. .8227)
/gene="elisa"
complement(7619. .8227)
/gene="elisa"
/note="putative urea/short-chain amide ABC transporter"
/codon_start=1
/transl_table=11
/product="ABC transporter-like protein"
/protein_id="AAG53594.1"
/db_xref="GI:12407619"
/translation="DCXLRTRXGNKLVSCIAGLLPFIASGTWVWHEAPGVNQAGTEI
XVYPQDRIRFSQLTVEENHLHQAQMAKGIAGSLSRDVFEFPPELYPLQLRASGLND
QQYQLALANALVTRPRLILLDPSRGTQQLFTVKLAQLLRLNRELGLSVLLAEQHLS
LIRRVADRECLLYRGRNVAQGRVSELDODPLIAHWMGPTASQ"
1512 a 2642 c 2605 g 1453 t 15 others
BASE COUNT

```





-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 13 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1140: contig of 1140 bp in length  
\* 1141 1240: gap of 100 bp  
\* 1241 1772: contig of 532 bp in length  
\* 1773 1872: gap of 100 bp  
\* 1873 6494: contig of 4622 bp in length  
\* 6495 6594: gap of 100 bp  
\* 6595 10712: contig of 4118 bp in length  
\* 10713 10812: gap of 100 bp  
\* 10813 15080: contig of 4268 bp in length  
\* 15081 15180: gap of 100 bp  
\* 15181 20098: contig of 4918 bp in length  
\* 20099 20198: gap of 100 bp  
\* 20199 25682: contig of 5484 bp in length  
\* 25683 25782: gap of 100 bp  
\* 25783 33225: contig of 7443 bp in length  
\* 33226 33325: gap of 100 bp  
\* 33326 40819: contig of 7494 bp in length  
\* 40820 40919: gap of 100 bp  
\* 40920 52456: contig of 11537 bp in length  
\* 52457 52556: gap of 100 bp  
\* 52557 70887: contig of 18331 bp in length  
\* 70888 70987: gap of 100 bp  
\* 70988 108187: contig of 37200 bp in length  
\* 108188 108287: gap of 100 bp  
\* 108288 155633: contig of 47346 bp in length.

## FEATURES

source  
1. .155633  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="4"  
/clone="RP11-673L23"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. .1140  
/note="assembly\_fragment"  
1241. .1772  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:left  
1873. .6494  
/note="assembly\_fragment"  
6595. .10712  
/note="assembly\_fragment"  
10813. .15080  
/note="assembly\_fragment"  
15181. .20098  
/note="assembly\_fragment"  
20199. .25682  
/note="assembly\_fragment"  
25783. .33225  
/note="assembly\_fragment"  
33326. .40819  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right  
40920. .52456  
/note="assembly\_fragment"  
52557. .70887  
/note="assembly\_fragment"  
70988. .108187  
/note="assembly\_fragment"  
108288. .155633  
/note="assembly\_fragment"  
50121 a 26755 c 26070 g 51486 t 1201 others

## ORIGIN

Query Match 80.8%; Score 20.2; DB 69; Length 155633;  
Best Local Similarity 88.0%; Pred. No. 23;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 attgtcgcaagtcacagaattattt 25  
||||| ||||| ||| |||||  
Db 4868 ATTGCTCAAGTAACATAATTATT 4892  
RESULT 19  
AC020741  
LOCUS AC020741 191804 bp DNA PRI 18-OCT-2000  
DEFINITION Homo sapiens clone RP11-798L4, complete sequence.  
ACCESSION AC020741  
VERSION AC020741.4 GI:10864254  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 191804)  
AUTHORS Waterston,R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 191804)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (08-JAN-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 3 (bases 1 to 191804)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (18-OCT-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
COMMENT On Oct 18, 2000 this sequence version replaced gi:7631118.  
Center project name: H\_NH0798L04.  
FEATURES  
source  
1. .191804  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-798L4"  
BASE COUNT 65994 a 34048 c 31965 g 59797 t  
ORIGIN  
Query Match 80.8%; Score 20.2; DB 87; Length 191804;  
Best Local Similarity 88.0%; Pred. No. 23;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 attgtcgcaagtcacagaattattt 25  
||||| ||||| ||| |||||  
Db 36094 ATTGCTCAAGTAACATAATTATT 36118  
RESULT 20  
AC084797  
LOCUS AC084797 172258 bp DNA HTG 17-NOV-2000  
DEFINITION Homo sapiens chromosome 16 clone RP11-6203, WORKING DRAFT SEQUENCE,  
31 unordered pieces.  
ACCESSION AC084797  
VERSION AC084797.1 GI:11192125  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 172258)  
AUTHORS DOE Joint Genome Institute.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Sequencing of Human Chromosome 16  
 2 (bases 1 to 172258)  
 DOE Joint Genome Institute.  
 Direct Submission  
 Submitted (17-NOV-2000) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 -----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: <http://www.jgi.doe.gov>  
 -----  
 Project Information  
 Center Project Name: 0  
 Center clone name: RPCI-11\_6203  
 -----  
 Summary Statistics  
 Consensus quality: 126064 bases at least Q40  
 Consensus quality: 142328 bases at least Q30  
 Consensus quality: 149310 bases at least Q20  
 Estimated insert size: 158300; agarose-fp estimation  
 Estimated insert size: 169258; sum-of-contigs estimation  
 Quality coverage: 4.23 in Q20 bases; agarose-fp estimation  
 Quality coverage: 3.95 in Q20 bases; sum-of-contigs estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 31 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1116: contig of 1116 bp in length  
 1117 1216: gap of unknown length  
 1217 2336: contig of 1140 bp in length  
 2337 2456: gap of unknown length  
 2457 3865: contig of 1409 bp in length  
 3866 3965: gap of unknown length  
 3966 5349: contig of 1384 bp in length  
 5350 5490: gap of unknown length  
 5491 6474: contig of 1025 bp in length  
 6475 6574: gap of unknown length  
 6575 7846: contig of 1272 bp in length  
 7847 9122: contig of 1176 bp in length  
 9123 9222: gap of unknown length  
 9223 10297: contig of 1074 bp in length  
 10297 10397: gap of unknown length  
 10397 11704: contig of 1308 bp in length  
 11705 11804: gap of unknown length  
 11805 13371: contig of 1567 bp in length  
 13372 13471: gap of unknown length  
 13472 15810: contig of 2339 bp in length  
 15811 15910: gap of unknown length  
 15911 18530: contig of 2620 bp in length  
 18531 18630: gap of unknown length  
 18631 21218: contig of 2588 bp in length  
 21219 21318: gap of unknown length  
 21319 24015: contig of 2697 bp in length  
 24016 24115: gap of unknown length  
 24116 27354: contig of 3239 bp in length  
 27355 27454: gap of unknown length  
 27455 29518: contig of 2064 bp in length  
 29519 29618: gap of unknown length  
 29619 33178: contig of 3560 bp in length  
 33179 33278: gap of unknown length  
 33279 40853: contig of 7575 bp in length  
 40854 40953: gap of unknown length  
 40954 45632: contig of 4678 bp in length  
 45632 45731: gap of unknown length  
 45732 48140: contig of 2409 bp in length  
 48141 48240: gap of unknown length  
 48241 54242: contig of 6002 bp in length

\* 54243 54342: gap of unknown length  
 \* 54343 62528: contig of 8186 bp in length  
 \* 62529 62628: gap of unknown length  
 \* 62629 69663: contig of 7035 bp in length  
 \* 69664 69763: gap of unknown length  
 \* 69764 79925: contig of 10162 bp in length  
 \* 79926 80025: gap of unknown length  
 \* 80026 91695: contig of 11670 bp in length  
 \* 91696 91795: gap of unknown length  
 \* 91796 98537: contig of 6742 bp in length  
 \* 98538 98637: gap of unknown length  
 \* 98638 107062: contig of 8425 bp in length  
 \* 107063 107162: gap of unknown length  
 \* 107163 119401: contig of 12239 bp in length  
 \* 119402 119501: gap of unknown length  
 \* 119502 132164: contig of 12663 bp in length  
 \* 132165 132264: gap of unknown length  
 \* 132265 150238: contig of 17974 bp in length  
 \* 150239 150338: gap of unknown length  
 \* 150339 172258: contig of 21920 bp in length.

FEATURES  
 Location/Qualifiers  
 1..172258  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /clone="RPCI1-6203"  
 /clone\_lib="RPCI human BAC library 11"  
 47622 a 36504 c 38345 g 45759 t 4028 others

BASE COUNT  
 ORIGIN

Query Match 77.6%; Score 19.4; DB 77; Length 172258;  
 Best Local Similarity 95.2%; Pred. No. 58;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 attgtcgcaagtccacgaatt 21  
 |||||  
 DB 58123 ATTGTGCGCAAGTGACAGATT 58143

RESULT 21  
 HS82J11 132981 bp DNA PRI 23-NOV-1999  
 LOCUS Human DNA sequence from PAC 82J11 and cosmid U134E6 on chromosome  
 DEFINITION Xq22. Contains NIK like and Thyroxin-binding globulin precursor  
 (T4-binding globulin, TBG) genes, ESTs and STSS.  
 ACCESSION Z83850  
 VERSION Z83850.1 GI:2780172  
 KEYWORDS globulin; NIK; T4; TBG; Thyroxin; Xq22.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 132981)  
 AUTHORS Ho,S.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-JAN-1998) Chromosome X Project Group  
 (http://www.sanger.ac.uk/HGP/chrX/) Sanger Centre, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquires:  
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Jan 16, 1998 this sequence version replaced gi:2467172.  
 IMPORTANT: This sequence is the entire insert of clone 82J11.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variations annotated may not be found in the sequence submission  
 corresponding to the overlapping clone as we submit sequences with  
 only a small overlap as described above.  
 This sequence was generated from part of bacterial clone contigs of  
 human chromosome X, constructed by the Sanger Centre chromosome X  
 mapping group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/chrX/  
 This sequence has been finished according to sequence map criteria

as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The true left end of clone 82J11 is at 1 in this sequence. The true left end of clone U134E6 is at 62897.

The true right end of clone U134E6 is at 102829.

The true right end of clone 82J11 is at 132981.

82J11 is from the library RPC11 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong.

For further details see <http://bacpac.med.buffalo.edu/U134E6> is from the Lawrence Livermore National Laboratory flow-sorted X chromosome cosmid library LLOXMC01.

## FEATURES

```

source
  1..132981
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
source
  1..92644
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="RP1-82J11"
    /clone_lib="RPC1-1"
    /size=1150..1340
    /note="L1R8 repeat: matches 561..383 of consensus"
    /note="11 copies of 2 mer 100 % conserved"
    /note="MER3 repeat: matches 208..1 of consensus"
    /note="L1 repeat: matches 5281..4409 of consensus"
    /note="L1R2 repeat: matches 34..893 of consensus"
    /note="L1R2 repeat: matches 34..893 of consensus"
    /note="21 copies of 2 mer 93 % conserved"
    /note="L1M4 repeat: matches 1039..66 of consensus"
    /note="L1P4 repeat: matches 890..825 of consensus"
    /note="L1 repeat: matches 5299..1333 of consensus"
    /note="L1 repeat: matches 2132..1556 of consensus"
    /note="L1M4 repeat: matches 931..333 of consensus"
    /note="MIR repeat: matches 4..254 of consensus"
    /note="AluJb repeat: matches 302..134 of consensus;
incomplete repeat"
    join(<25698..25832,27946..28089,30381..30521,34448..34548,
36822..36971,40106..40241,43541..43680,46012..>46099)
    /gene="dJ82J11.1"
    /note="mouse NIK serine threonine protein kinase like;
match ESTs R98571 RA333635"
    /evidence=not_experimental
    25698..46099
    /gene="dJ82J11.1"
    join(25698..25832,27946..28089,30381..30521,34448..34548,
36822..36971,40106..40241,43541..43680)
    /partial
    /gene="dJ82J11.1"
    /note="mouse NIK serine threonine protein kinase like;
match: proteins p97820 CE02384"
    /codon_start=3
    /evidence=not_experimental
    /protein_id="CA806091.1"
    /db_xref="GI:2780173"
    /translation="PEESKPQSEVNVNPLVSPACKPLIHMYKEFTSETCCGSLW
GVNLLGTRNLVMDRSGKADITKLIRPPFRQIQVLEPLNLTISGHNKRLRYVH
LTWLNKILNDDPKRRROEMLKTEACAKIDKLTGFOHEETYIAIALKSIHLIYA
WAKPSESTAIKVFPLDHPVTDLAIGSEKRLKIFFSSADGYHLIDAESVMSDV
TLPKNIIILIDCLGIGOMLTFNAEALSVEANQLPKKILLEMWKDIPSSIAFECTORT

```

```

repeat_region
  26070..26296
    /note="AluSg repeat: matches 51..279 of consensus;
incomplete repeat"
    29261..29598
    /note="THE1C repeat: matches 371..36 of consensus"
    30655..30695
    /note="5S repeat: matches 41..1 of consensus"
    32254..33127
    /note="L1MB8 repeat: matches 920..1 of consensus"
    33130..33431
    /note="AluSx repeat: matches 296..1 of consensus"
    33435..34221
    /note="L1 repeat: matches 5248..4454 of consensus"
    34694..34898
    /note="AluJo repeat: matches 301..100 of consensus;
incomplete repeat"
    35592..35734
    /note="MER5A repeat: matches 189..43 of consensus"
    37678..37969
    /note="AluYa5 repeat: matches 10..301 of consensus"
    41035..41084
    /note="25 copies of 2 mer 100 % conserved"
    41321..41354
    /note="17 copies of 2 mer 91 % conserved"
    45370..45533
    /note="L1MA8 repeat: matches 862..1031 of consensus"
    46370..49114
    /note="match: ESTs N30117 M60729 N33939 R98572 R28658
R70758 R62577 H47233 R62576 H06699 R70808 R28510
H06749 AA331981 AA335043 AA249339 AA247203 H03220
AA331981"
    complement(47194..47616)
    /note="match STRS G28348 G26027"
    complement(49515..49722)
    /note="match STRS AF020167"
    50129..50131
    50194..50574
    /note="L1 repeat: matches 7..388 of consensus"
    50573..55498
    /note="L1 repeat: matches 479..5390 of consensus"
    55353..56235
    /note="L1PA2 repeat: matches 1..891 of consensus"
    57119..57154
    /note="18 copies of 2 mer 81 % conserved"
    58395..58799
    /note="MLT1A1 repeat: matches 362..4 of consensus"
    59693..59994
    /note="AluSx repeat: matches 302..1 of consensus"
    59837..59917
    60007..60053
    /note="MST repeat: matches 48..1 of consensus"
    60065..61105
    /note="MST-INTERNAL repeat: matches 1621..628 of
consensus"
    61117..61960
    /note="L1 repeat: matches 4530..5388 of consensus"
    61829..62702
    /note="L1PA7 repeat: matches 1..887 of consensus"
    62721..62912
    /note="MST-INTERNAL repeat: matches 620..433 of consensus"
    63677..63679
    /note="clone U134P6; GGA in this entry; insertion"
    /replace="ga"
    64668..64889
    /note="L1ME3A repeat: matches 349..565 of consensus"
    65224..65456
    /note="L1ME3A repeat: matches 910..691 of consensus"
    65303..65456
    /note="L1ME3 repeat: matches 839..691 of consensus"
    66620..66721
    /note="MER20 repeat: matches 23..125 of consensus"
    66973..67866

```

```

repeat_region /note="L1PA15 repeat: matches 897. .1 of consensus"
67723. .72165
repeat_region /note="L1 repeat: matches 5390. .987 of consensus"
73816. .74035
repeat_region /note="MIR repeat: matches 249. .18 of consensus"
74040. .74096
repeat_region /note="L1MC1 repeat: matches 1079. .1022 of consensus"
74155. .74188
repeat_region /note="17 copies of 2 mer 82 & conserved"
74288. .74681
repeat_region /note="M1RL1A1 repeat: matches 362. .2 of consensus"
76306. .76443
repeat_region /note="MIR2 repeat: matches 138. .1 of consensus"
77147. .77434
repeat_region /note="AluJo repeat: matches 16. .302 of consensus;
incomplete repeat"
82377. .78947
repeat_region /note="L1 repeat: matches 3606. .2975 of consensus"
79099. .79390
repeat_region /note="AluX repeat: matches 302. .1 of consensus"
79413. .79776
repeat_region /note="L1 repeat: matches 2512. .2123 of consensus"
79882. .80048

```

```

Query Match 76.88; Score 19.2; DB 92; Length 132981;
Best Local Similarity 87.5%; Pred. No. 73;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```
Oy 2 ttgtcgcaagtcacagaattatt 25
```

```
Db 84233 TTGTCAAAATGACAGAAATATT 84256
```

```

RESULT 22
AC020546/c
LOCUS AC020546 153087 bp DNA HTG 17-AUG-2000
DEFINITION Homo sapiens chromosome 14 clone RP11-95F16, WORKING DRAFT
SEQUENCE, 36 unordered pieces.
ACCESSION AC020546
VERSION AC020546.2 GI:9502458
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston,R.H.
1 (bases 1 to 153087)
The sequence of Homo sapiens clone
Waterston,R.H.
Unpublished
2 (bases 1 to 153087)
Direct Submission
Waterston,R.H.
Submitted (03-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jul 27, 2000 this sequence version replaced gi:6665582.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0095F16
----- Summary Statistics -----
Sequencing vector: M13; 93%
Chemistry: Dye-terminator; 7%
Assembly: Dye-terminator Big Dye; 7% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 133287 bases at least Q40
Consensus quality: 139089 bases at least Q30
Consensus quality: 142441 bases at least Q20
Insert size: 151000; agarose-fp

```

```

Insert size: 149587; sum-of-contigs
Quality coverage: 3.21 in Q20 bases; agarose-fp
Quality coverage: 3.10 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

* 1 1284: contig of 1284 bp in length
* 1285 1384: gap of unknown length
* 1385 1385: contig of 1872 bp in length
* 3257 3356: gap of unknown length
* 3357 3557: contig of 1805 bp in length
* 5162 5261: gap of unknown length
* 5262 5262: contig of 2461 bp in length
* 7723 7822: gap of unknown length
* 7823 7823: contig of 1985 bp in length
* 9808 9907: gap of unknown length
* 9908 11659: contig of 1752 bp in length
* 11660 11759: gap of unknown length
* 11760 14961: contig of 3202 bp in length
* 14962 15061: gap of unknown length
* 15062 17341: contig of 2280 bp in length
* 17342 17441: gap of unknown length
* 17442 20613: contig of 3172 bp in length
* 20614 20713: gap of unknown length
* 20714 24483: contig of 3770 bp in length
* 24484 24583: gap of unknown length
* 24584 27629: contig of 3046 bp in length
* 27630 27729: gap of unknown length
* 27730 29886: contig of 2257 bp in length
* 29887 30086: gap of unknown length
* 30087 32919: contig of 2833 bp in length
* 32920 33019: gap of unknown length
* 33020 35697: contig of 2678 bp in length
* 35698 35797: gap of unknown length
* 35798 38277: contig of 2480 bp in length
* 38278 38377: gap of unknown length
* 38378 41653: contig of 3276 bp in length
* 41654 41753: gap of unknown length
* 41754 44674: contig of 2921 bp in length
* 44675 44774: gap of unknown length
* 44775 49609: contig of 4835 bp in length
* 49610 49709: gap of unknown length
* 49710 53789: contig of 4080 bp in length
* 53790 53889: gap of unknown length
* 53890 57454: gap of unknown length
* 57454 61602: contig of 4049 bp in length
* 61603 61702: gap of unknown length
* 61703 65138: contig of 3436 bp in length
* 65139 65238: gap of unknown length
* 65239 70266: contig of 5028 bp in length
* 70267 70366: gap of unknown length
* 70367 75388: contig of 5022 bp in length
* 75389 75489: gap of unknown length
* 75489 80195: contig of 4707 bp in length
* 80196 80295: gap of unknown length
* 80296 85960: contig of 5665 bp in length
* 85961 86060: gap of unknown length
* 86061 91136: contig of 5076 bp in length
* 91137 91236: gap of unknown length
* 91237 97282: contig of 6046 bp in length
* 97283 104072: gap of unknown length
* 104073 104172: gap of unknown length
* 104173 111143: contig of 6971 bp in length
* 111144 111243: gap of unknown length
* 111244 118031: contig of 6788 bp in length

```

```

* 118032 118131: gap of unknown length
* 118132 122562: contig of 4431 bp in length
* 122562 122662: gap of unknown length
* 122662 128512: contig of 5850 bp in length
* 128512 128613: gap of unknown length
* 128613 135250: contig of 6638 bp in length
* 135250 135351: gap of unknown length
* 135351 144541: gap of unknown length
* 144541 144641: contig of 9190 bp in length
* 144641 153087: contig of 8447 bp in length.

```

## FEATURES

## source

```

1. 153087
  Location/Qualifiers
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="14"
    /clone="RP11-95F16"

```

```
1. 1284
```

```

misc_feature /note="assembly_name:Contig9"
misc_feature 1385..3256 /note="assembly_name:Contig10"
misc_feature 3357..5161 /note="assembly_name:Contig11"
misc_feature 5262..7722 /note="assembly_name:Contig12"
misc_feature 7823..9807 /note="assembly_name:Contig13"
misc_feature 9908..11659 /note="assembly_name:Contig14"
misc_feature 11760..14961 /note="assembly_name:Contig15"
misc_feature 15062..17341 /note="assembly_name:Contig16"
misc_feature 17442..20613 /note="assembly_name:Contig17"
misc_feature 20714..24483 /note="assembly_name:Contig18"
misc_feature 24584..27629 /note="assembly_name:Contig19"
misc_feature 27730..29986 /note="assembly_name:Contig20"
misc_feature 30087..32919 /note="assembly_name:Contig21"
misc_feature 33020..35697 /note="assembly_name:Contig22"
misc_feature 35798..38277 /note="assembly_name:Contig23"
  clone_end:T7
  vector_side:left
misc_feature 38378..41653 /note="assembly_name:Contig24"
misc_feature 41734..44674 /note="assembly_name:Contig25"
misc_feature 44775..49609 /note="assembly_name:Contig26"
misc_feature 49710..53789 /note="assembly_name:Contig27"
misc_feature 53890..57453 /note="assembly_name:Contig28"
misc_feature 57534..61602 /note="assembly_name:Contig29"
misc_feature 61703..65138 /note="assembly_name:Contig30"
misc_feature 65239..70266 /note="assembly_name:Contig31"
misc_feature 70367..75388 /note="assembly_name:Contig32"
misc_feature 75489..80195 /note="assembly_name:Contig33"
misc_feature 80296..85960 /note="assembly_name:Contig34"
misc_feature 86061..91136 /note="assembly_name:Contig35"
misc_feature 91237..97282

```

```

/note="assembly_name:Contig36"
97383..104072
/note="assembly_name:Contig37"
104173..111143
/note="assembly_name:Contig38"
111244..118031
/note="assembly_name:Contig39"
118132..122562
/note="assembly_name:Contig40"
122663..128512
/note="assembly_name:Contig41"
128613..135250
/note="assembly_name:Contig42"
135351..144540
/note="assembly_name:Contig43"
144641..153087

```

```

Query Match 76.8%; Score 19.2; DB 65; Length 153087;
Best Local Similarity 87.5%; Pred. No.73;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```
QY 2 ttgtcgcaagtcacagaattattt 25
```

```
DB 116175 TTCTCAAGTCACAGATCAATT 116152
```

## RESULT 23

## CNS05TC0

```

LOCUS Homo sapiens chromosome 14 clone R-15E14, ** SEQUENCING IN
DEFINITION PROGRESS ***, in ordered pieces.
ACCESSION AL352981
VERSION AL352981.3 GI:10241556
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.

```

## ORGANISM

```

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

```

Direct Submission
Submitted (19-SEP-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
On Sep 20, 2000 this sequence version replaced gi:9213044.
----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: Seqref@genoscope.cns.fr
-----

```

IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage, etc. . . even if efforts are made to eliminate these contaminating sequences. The following BAC sequence is oriented from the T7 to the SP6 end.

Upstream BAC (overlapping the T7 end) : R-74H1  
Downstream BAC (overlapping the SP6 end) : R-6101

## Overall quality chart :

```

Range : bases
0 : 1
1 - 9 : 33
10 - 19 : 316
20 - 29 : 941
30 - 39 : 3472
40 - 49 : 13463
50 - 59 : 14709
60 - 69 : 8224
70 - 79 : 14329

```

80 - 89 : 36335  
90 - 99 : 66390

Percentage of bases with a quality value >= 40 : 96 %.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

# FEATURES

source  
1. .158213  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/clone="R-15E14"  
/clone\_lib="RPCI-11"

BASE COUNT 45522 a 33565 c 33201 g 45924 t 1 others  
ORIGIN

Query Match 76.8%; Score 19.2; DB 84; Length 158213;  
Best Local Similarity 87.5%; Pred. No. 73;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgtcgaagtcacagaattattt 25  
|||||  
Db 43079 TTCTCACAAGTCACAGATCAATT 43102

# RESULT 24

AC022047  
LOCUS  
DEFINITION Homo sapiens clone Rp11-15E14, WORKING DRAFT SEQUENCE, 5 unordered  
pieces

AC022047  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

# REFERENCE

AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 161433)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens, clone Rp11-15E14  
Unpublished  
2 (bases 1 to 161433)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,  
Boguslavsky,L., Bouknight,B., Brown,A., Burkett,G., Castle,A.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,  
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K.,  
Macdonald,P., Marquitz,J., McEwan,P., McGurk,A., McKernan,K.,  
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,  
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Tirell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and Zody,M.

# TITLE

JOURNAL  
Submitted (25-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Feb 28, 2000 this sequence version replaced gi:6850470.  
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WITB  
Web site: http://www.seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information

Center project name: L3449  
Center clone name: 15\_E14

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 158759 bases at least Q40  
Consensus quality: 159371 bases at least Q30  
Consensus quality: 159762 bases at least Q20  
Insert size: 157000; agarose-fp  
Quality coverage: 9.8 in Q20 bases; agarose-fp  
Quality coverage: 9.5 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1153: contig of 1153 bp in length  
\* 1154 1253: gap of 100 bp  
\* 1254 4645: contig of 3392 bp in length  
\* 4646 4745: gap of 100 bp  
\* 4746 45083: contig of 40338 bp in length  
\* 45084 45183: gap of 100 bp  
\* 45184 91139: contig of 45956 bp in length  
\* 91140 91239: gap of 100 bp  
\* 91240 161433: contig of 70194 bp in length.

# FEATURES

source  
1. .161433  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Rp11-15E14"  
/clone\_lib="RPCI-11 Human Male BAC"

# misc\_feature

1. .1153  
/note="assembly\_fragment"  
1254. 4645  
/note="assembly\_fragment"

# misc\_feature

vector\_side:right  
4746. 45083  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right

# misc\_feature

45184. 91139  
/note="assembly\_fragment"  
91240. 161433  
/note="assembly\_fragment"

BASE COUNT 46681 a 33931 c 34009 g 46410 t 402 others  
ORIGIN

Query Match 76.8%; Score 19.2; DB 67; Length 161433;  
Best Local Similarity 87.5%; Pred. No. 73;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgtcgaagtcacagaattattt 25  
|||||

Db 85492 TTCTCACAAGTCACAGATCAATT 85515

# RESULT 25

LOCUS  
DEFINITION Homo sapiens chromosome 14 clone R-74H1, \*\*\* SEQUENCING IN PROGRESS  
ACCESSION AL445594  
VERSION AL445594.2 GI:11967629

CNS07ED4 171409 bp DNA HTG 21-DEC-2000  
Homo sapiens chromosome 14 clone R-74H1, \*\*\* SEQUENCING IN PROGRESS

```

KEYWORDS      HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE        human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 171409)
              Direct Submission
              Genoscope.
              Submitted (19-DEC-2000) Genoscope - Centre National de Sequencage :
              BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
              Web : www.genoscope.cns.fr)
              On Dec 22, 2000 this sequence version replaced gi:10880262.
              ----- Genome Center
              Center: Genoscope / Centre National de Sequencage
              Center code: GS
              Web site: http://www.genoscope.cns.fr/
              Contact: SeqRef@genoscope.cns.fr
              -----
              IMPORTANT: This sequence is unfinished and does not necessarily
              represent the correct sequence. Work on the sequence is in progress
              and the release of this data is based on the understanding that the
              sequence may change as work continue. The sequence may be
              contaminated with foreign sequence from E.coli, yeast, vector,
              phage, etc. . . even if efforts are made to eliminate these
              contaminating sequences. The following BAC sequence is oriented
              from the T7 to the SP6 end.
              Upstream BAC (overlapping the T7 end) : R-20AN11
              Downstream BAC (overlapping the SP6 end) : R-15E14 (AC-AL352981)
              ----- Summary Statistics
              Assembly program: Phrap; version 2.0
              Quality coverage: 7.19x in Q20 bases; sum-of-contigs
              -----
              Overall quality chart :
              Range : bases
              0 : 345
              1 - 9 : 36
              10 - 19 : 359
              20 - 29 : 881
              30 - 39 : 2494
              40 - 49 : 11222
              50 - 59 : 12150
              60 - 69 : 12390
              70 - 79 : 25439
              80 - 89 : 51258
              90 - 99 : 54835
              -----
              Percentage of bases with a quality value >= 40 : 97 %.
              * NOTE: This is a 'working draft' sequence.
              * This sequence will be replaced
              * by the finished sequence as soon as it is available and
              * the accession number will be preserved.
              Location/Qualifiers
                1. 171409
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="14"
                  /clone="R-74H1"
                  /clone_lib="RPC1-11"
              BASE COUNT 49330 a 36078 c 36667 g 49332 t      2 others
              ORIGIN

              Query Match      76.8%; Score 19.2; DB 84; Length 171409;
              Best Local Similarity 87.5%; Pred. No. 73;
              Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

              QY      2 ttgtcgaagtccacagaattattt 25
              Db 124876 TTCTCACAAGTCACAGATCAATTT 124899

              RESULT 26

```

AC008118/c  
LOCUS  
DEFINITION

AC008118 180015 bp DNA HTG 17-APR-2001  
Homo sapiens chromosome 12 clone RP11-515D8, WORKING DRAFT  
SEQUENCE, 6 unordered pieces.

AC008118  
AC008118.18 GI:13654317  
HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTDP.  
human.

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 180015)

Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbarella,J.,  
Benton,J., Binnage,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Carron,T.F.,  
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,  
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,  
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,J., Jolivet,S.,  
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,  
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,  
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
Louissegue,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,  
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,  
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokoko,S.,  
Ogih,M., Okunou,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,  
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,  
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,  
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,  
Stone,H., Sutton,A., Svatek,A., Tabor,P., Telford,B., Thomas,N.,  
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,  
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,  
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,  
Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,  
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 180015)

Worley,K.C.

Direct Submission

Submitted (24-JUL-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Apr 17, 2001 this sequence version replaced gi:13592160.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: HMIH

Center clone name: RP11-515D8

----- Summary Statistics

Sequencing vector: Plasmid; M77789

Sequencing vector: M13; L08821

Chemistry: Dye-primer Bodypy; 14% of reads

Chemistry: Dye-terminator Big Dye; 86% of reads

Assembly program: Phrap; version 0.990329  
 Consensus quality: 176244 bases at least Q40  
 Consensus quality: 178377 bases at least Q30  
 Consensus quality: 179104 bases at least Q20  
 Estimated insert size: 178510; sum-of-contigs estimation  
 Estimated insert size: 162000; agarose-fp estimation  
 Quality coverage: 7.7x in Q20 bases; agarose-fp estimation  
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 6 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 62796: contig of 62796 bp in length  
 62897 62896: gap of unknown length  
 62897 125066: contig of 62170 bp in length  
 125067 125166: gap of unknown length  
 125167 153355: contig of 28189 bp in length  
 153356 153455: gap of unknown length  
 153456 168583: contig of 15128 bp in length  
 168584 168583: gap of unknown length  
 168684 175046: contig of 6363 bp in length  
 175047 175146: gap of unknown length  
 175147 180015: contig of 4869 bp in length.

## FEATURES

source

1. 180015  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="12"  
 /clone="RP11-685B14"

BASE COUNT 56442 a 33634 c 33364 g 54070 t 505 others  
 ORIGIN

Query Match 76.8%; Score 19.2; DB 60; Length 180015;  
 Best Local Similarity 87.5%; Pred. No. 73;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattatt 24  
 ||| ||||| ||||| ||||| |||||

Db 12902 ATTTAGCAAGTCAGAAATATT 12879

RESULT 27  
 AC025062  
 LOCUS AC025062 190539 bp DNA HTG 03-MAR-2001  
 DEFINITION Homo sapiens chromosome 8 clone RP11-685B14 map 8, WORKING DRAFT  
 SEQUENCE, 2 ordered pieces.  
 ACCESSION AC025062  
 VERSION AC025062.3 GI:13184200  
 KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 190539)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 TITLE Homo sapiens chromosome 8, clone RP11-685B14  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 190539)  
 AUTHORS Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
 Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,  
 Campiano,A., Cooke,P., Choepel,Y., Colangelo,M., Collins,S.,  
 Collymore,A., Cooley,A., DeArelano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., Lakocque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
 Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,  
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
 Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

## TITLE

JOURNAL

## COMMENT

Submitted (04-MAR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Mar 2, 2001 this sequence version replaced gi:17684472.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7473

Center clone name: 685\_B14

----- Summary Statistics

Sequencing vector: M13; M77815; 44% of reads  
 Sequencing vector: Plasmid; n/a; 56% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 190047 bases at least Q40  
 Consensus quality: 190291 bases at least Q30  
 Consensus quality: 190380 bases at least Q20  
 Insert size: 192000; agarose-fp  
 Insert size: 190439; sum-of-contigs  
 Quality coverage: 7.9 in Q20 bases; agarose-fp  
 Quality coverage: 7.9 in Q20 b.

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.

\* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

\* 1 176380: contig of 176380 bp in length  
 \* 176381 176480: gap of 100 bp  
 \* 176481 190539: contig of 14059 bp in length.

## FEATURES

source

1. 190539  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="8"  
 /map="8"

/clone="RP11-685B14"  
 /clone.lib="RP11-11 Human Male BAC"  
 1. 176380

## misc\_feature

/note="assembly\_fragment  
 vector\_side:left"  
 176481..190539

## misc\_feature

/note="assembly\_fragment  
 clone\_end:T7  
 vector\_side:right"

## BASE COUNT

ORIGIN

61261 a 37428 c 36378 g 55372 t 100 others



Query Match 76.8%; Score 19.2; DB 69; Length 190539;  
 Best Local Similarity 87.5%; Pred. NO. 73;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattattt 25  
 ||||| ||||| ||||| ||||| ||  
 Db 111817 TTGTGCAAGTCACAAAATTACTT 111840

# RESULT 28 AC007564/c

LOCUS AC007564 194058 bp DNA PRI 03-JUL-1999  
 DEFINITION Homo sapiens 12q22 BAC RPC111-513P18 (Roswell Park Cancer Institute  
 Human BAC Library) complete sequence.

ACCESSION AC007564  
 VERSION AC007564.9 GI:5306220

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## AUTHORS

Muzny,D., Aronson,A.D., Bouck,J., Bunac,C., Chen,Z., Ding,Y.,  
 Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H.,  
 Gorrell,L.L., Hernandez,J., Issar,A., Jackson,L., Kneitz,S.,  
 Kondejewski,N., Lau,S., Leal,B., Lee,E., Lichtarge,O., Liu,W.,  
 Logan,O., Lu,J., Marandel,I., Martinez,C., Merscher,S., Miller,A.,  
 Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L.,  
 Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shim,C., Simon,M.,  
 Vo,O., Williamson,A., Worley,K.C., Xhang,A.M., Yang,R., Yu,W.,  
 Zhou,X., Kucherlapati,R., Nelson,D. and Gibbs,R.A.

## TITLE

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 194058)

AUTHORS Worley,K.C.

## TITLE

JOURNAL Submitted (15-MAY-1999) Molecular and Human Genetics, Baylor  
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 194058)

AUTHORS Worley,K.C.

## TITLE

JOURNAL Submitted (01-JUL-1999) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE 4 (bases 1 to 194058)

AUTHORS Worley,K.C.

## TITLE

JOURNAL Submitted (03-JUL-1999) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

On Jul 1, 1999 this sequence version replaced gi:5263308.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-help@bcm.tmc.edu

## COMMENT

CLONE LENGTH: This sequence does not necessarily represent the  
 entire insert of this clone. Overlapping regions of clones are only  
 sequenced and submitted once, so the sequence for the remainder of  
 the insert may be found in the record for the adjacent clones.  
 Overlapping clones are noted at the beginning and end of the  
 Features listing.

## ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
 of a local database that includes entries from dbSTS, GDB, and  
 local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
 unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
 EST and cDNA sequences. Genes demonstrate at least two exons  
 flanked by consensus splice sites that maintained sequence  
 continuity across the splice junctions. Sequences that are not  
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
 standard of double strand coverage with a minimum of 2 clones and 2  
 reads with no ambiguities or 2 chemistries with a minimum of 2  
 clones and 3 reads with no ambiguities. If the sequence quality for  
 a region does not meet this standard, it will be indicated in the  
 annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
 standards - estimated error rate less than 1 per 10,000 bases.  
 Reports of lowest quality individual bases and measures of base  
 quality are listed below. Description of the metrics can be found  
 at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT-----

## ----- Summary Statistics -----

Contig length: 122887

Phrap values in estimate: 121646

Average error rate (BCM-Phrap estimate): 0.000150138

Fraction of Phrap values less than 40 : 0.0216119

Number of consensus changing edits: 45

Number of N's in consensus : 0

## ----- Consensus changing edits -----

Position	Original+Context	Edited+Context
66363	acataatgtgt(n)tgngtatata	acataatgtgt(g)tgngtatata
66366	tatgtgtntg(n)gtatatatat	tatgtgtgt(g)gtatatatat
66426	tatacacaca(n)acacacacat	tatacacaca(c)acacacacat
71322	cgtcacact(n)taaaagctgt	cgtcacact(t)taaaagctgt
71375	ttcactgtct(n)taaaatgaat	ttcactgtct(t)taaaatgaat
71414	ggtataagat(n)acaaaaaat	ggtataagat(t)acaaaaaat
78081	tgacttttt(n)ttctttcagc	tgacttttt(t)ttctttcagc
78101	cactttaag(n)ngncanncca	cactttaag(a)aggaaccca
78102	actttaag(n)ngncanncca	actttaaga(g)gaaccca
78104	tittaagng(n)caannccat	tittaagng(g)caannccat
78108	agngnncaa(n)ncattatcac	agngnncaa(c)ncattatcac
78109	agngnncaa(n)ncattatcac	agngnncaa(c)ncattatcac
78119	ncattatcac(n)tggtctccaa	ncattatcac(t)tggtctccaa
78194	gtgtttttc(n)tggtgtgat	gtgtttttc(t)tggtgtgat
78224	tttctctat(n)tttgatttcc	tttctctat(t)tttgatttcc
78272	gtgtttttc(n)tttatttacc	gtgtttttc(t)tttatttacc
78989	tattttgtct(n)tgctgttagc	tattttgtct(t)tgctgttagc
102112	acatgtgtga(n)ncgtctctac	acatgtgtga(a)ncgtctctac
102113	catgtgtgaac(n)cgctctact	catgtgtgaac(c)cgctctact
102125	gtctctacta(n)taatacaata	gtctctacta(t)taatacaata
102126	ttctctacta(n)taatacaata	ttctctacta(t)taatacaata
102134	attaatacaa(n)taatacaata	attaatacaa(a)taatacaata
102171	ctactgtgct(n)taatacaata	ctactgtgct(a)taatacaata
102183	ctactgtgct(n)taatacaata	ctactgtgct(a)taatacaata
102207	aatcactcaa(n)ctgtgagagt	aatcactcaa(a)ctgtgagagt
102225	ggtagaggtc(n)caatgagcca	ggtagaggtc(a)caatgagcca
102241	agccaagatt(n)tgcatgtga	agccaagatt(g)tgcatgtga
102808	gagttgtgt(n)caataaaaaa	gagttgtgt(t)caataaaaaa
102851	gtgtctgtgc(n)caatgagntc	gtgtctgtgc(a)caatgagntc
102857	tggtcagag(n)tgnttaaaaa	tggtcagag(c)tgnttaaaaa
102860	ccagagagntc(n)taaaacagt	ccagagagntc(c)taaaacagt
102883	aactgacat(n)ttttatatac	aactgacat(c)ttttatatac
102933	taaccaatag(n)tgatgttta	taaccaatag(c)tgatgttta
102939	atagntgatg(n)tttaatacat	atagntgatg(a)tttaatacat
103016	tcgggagatg(n)gcacaagtat	tcgggagatg(t)gcacaagtat
103085	cagctcacca(n)gagagtgtg	cagctcacca(g)gagagtgtg
103108	aggagaggtc(n)nnccanagga	aggagaggtc(a)nnccanagga
103109	ggagaggtc(n)nnccanagga	ggagaggtc(a)nnccanagga
103110	ggagaggtc(n)nnccanagga	ggagaggtc(a)nnccanagga
103114	ggcannncc(n)agggagagntc	ggcannncc(c)agggagagntc
103122	accagagag(n)gtcggagag	accagagag(t)gtcggagag
103129	acctccctac(n)aaacataaat	acctccctac(c)aaacataaat
122536	tataataat(n)atataatata	tataataat(t)atataatata
122825	ataataactgc(n)cactatatat	ataataactgc(a)cactatatat

122854	ataactgcac(n)ctatatataa	ataactgcac(a)ctatatataa
----- Distribution of Quality < 40 Bases -----		
	10001	*
	9001	*
	8001	*
	7001	*
#	6001	*
bases	5001	*
	4001	*
	3001	*
	2001	*
	1001	*
	01	*
-----		
Version: 1.01 qxf.		
FEATURES	Location/Qualifiers	
Source	1. 194058	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="RPC111-513P18"	
	/chromosome="12q22"	
gene	complement(join(2490..2646,3544..3700,4577..4603))	
	/gene="zu10h04.r1 Homo sapiens cDNA, AA470029"	
repeat_region	3702..3745	
	/rpt_family="AT_rich"	
repeat_region	5141..5666	
	/rpt_family="MT1F"	
repeat_region	6057..6201	
	/rpt_family="MIR"	
repeat_region	complement(9269..9311)	
	/rpt_family="L2"	
repeat_region	complement(9283..9451)	
	/rpt_family="MIR"	
repeat_region	9503..9542	
	/rpt_family="(GA)n"	
repeat_region	9544..9665	
	/rpt_family="L1MC4"	
repeat_region	12646..12787	
	/rpt_family="MER20"	
repeat_region	13503..13803	
	/rpt_family="AluX"	
repeat_region	13826..13866	
	/rpt_family="(GAAA)n"	
repeat_region	15405..15510	
	/rpt_family="MIR"	
repeat_region	complement(15799..15875)	
	/rpt_family="(TA)n"	
misc_feature	15808..15898	
	/function="Low coverage."	
repeat_region	complement(15882..16014)	
	/rpt_family="FLAM_C"	
repeat_region	complement(16018..16049)	
	/rpt_family="AT_rich"	
repeat_region	16298..16390	
	/rpt_family="MER58A"	
repeat_region	16407..16438	
	/rpt_family="AT_rich"	
Query Match		
Best Local Similarity	76.8%; Score 19.2; DB 86; Length 194058;	
Matches	21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
Qy	1 attgtcgaagtcacagaattatt 24	
Db 188503	ATTTAGCAAGTCATAGAAATTATT 188480	

RESULT 29  
AC026977  
LOCUS  
DEFINITION  
SEQUENCE, 51 unordered pieces.  
AC026977  
AC026977.2 GI:7677902  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 194718)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 18, clone RP11-768B23  
Unpublished  
2 (bases 1 to 194718)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,  
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Horton,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Kann,L., Karatas,A.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Lehotzky,J.,  
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehotzky,J.,  
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mieng,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,K., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Tagilillo,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (27-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 2, 2000 this sequence version replaced gi:7330323.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L8654  
Center clone name: 768\_B\_23  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 171617 bases at least Q40  
Consensus quality: 182763 bases at least Q30  
Consensus quality: 187263 bases at least Q20  
Insert size: 210000; agarose-fp  
Insert size: 189718; sum-of-contigs  
Quality coverage: 3.2 in Q20 bases; agarose-fp  
Quality coverage: 3.5 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 51 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence

AC026977 194718 bp DNA HTG 02-MAY-2000  
Homo sapiens chromosome 18 clone RP11-768B23 map 18, WORKING DRAFT  
SEQUENCE, 51 unordered pieces.  
AC026977  
AC026977.2 GI:7677902  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 194718)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 18, clone RP11-768B23  
Unpublished  
2 (bases 1 to 194718)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,  
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Horton,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Kann,L., Karatas,A.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Lehotzky,J.,  
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehotzky,J.,  
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mieng,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,K., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Tagilillo,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (27-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 2, 2000 this sequence version replaced gi:7330323.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L8654  
Center clone name: 768\_B\_23  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 171617 bases at least Q40  
Consensus quality: 182763 bases at least Q30  
Consensus quality: 187263 bases at least Q20  
Insert size: 210000; agarose-fp  
Insert size: 189718; sum-of-contigs  
Quality coverage: 3.2 in Q20 bases; agarose-fp  
Quality coverage: 3.5 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 51 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will  
\* be preserved.

```

* 1164: contig of 1164 bp in length
* 1165 1264: gap of 100 bp
* 1265 2302: contig of 1038 bp in length
* 2303 2402: gap of 100 bp
* 2403 3517: contig of 1115 bp in length
* 3518 3617: gap of 100 bp
* 3618 4687: contig of 1070 bp in length
* 4688 4787: gap of 100 bp
* 4788 6355: contig of 1568 bp in length
* 6356 6455: gap of 100 bp
* 6456 7718: contig of 1263 bp in length
* 7719 7818: gap of 100 bp
* 7819 9644: contig of 1826 bp in length
* 9645 9744: gap of 100 bp
* 9745 11368: contig of 1624 bp in length
* 11369 11468: gap of 100 bp
* 11469 13195: contig of 1727 bp in length
* 13196 13295: gap of 100 bp
* 13296 14769: contig of 1474 bp in length
* 14770 14869: gap of 100 bp
* 14870 15993: contig of 1124 bp in length
* 15994 16093: gap of 100 bp
* 16094 17326: contig of 1233 bp in length
* 17327 17426: gap of 100 bp
* 17427 18992: contig of 1566 bp in length
* 18993 19092: gap of 100 bp
* 19093 20718: contig of 1626 bp in length
* 20719 20818: gap of 100 bp
* 20819 22360: contig of 1542 bp in length
* 22361 22460: gap of 100 bp
* 22461 23830: contig of 1370 bp in length
* 23831 23930: gap of 100 bp
* 23931 25375: contig of 1445 bp in length
* 25376 25475: gap of 100 bp
* 25476 27805: contig of 2330 bp in length
* 27806 27905: gap of 100 bp
* 27906 30221: contig of 2316 bp in length
* 30222 30321: gap of 100 bp
* 30322 32276: contig of 1955 bp in length
* 32277 32376: gap of 100 bp
* 32377 35222: contig of 2846 bp in length
* 35223 35322: gap of 100 bp
* 35323 38008: contig of 2686 bp in length
* 38009 38108: gap of 100 bp
* 38109 40419: contig of 2311 bp in length
* 40420 40519: gap of 100 bp
* 40520 43373: contig of 2854 bp in length
* 43374 43473: gap of 100 bp
* 43474 45117: contig of 1644 bp in length
* 45118 45217: gap of 100 bp
* 45218 48518: contig of 3301 bp in length
* 48519 48618: gap of 100 bp
* 48619 51358: contig of 2740 bp in length
* 51359 51458: gap of 100 bp
* 51459 54472: contig of 3014 bp in length
* 54473 54572: gap of 100 bp
* 54573 57288: contig of 2716 bp in length
* 57289 57388: gap of 100 bp
* 57389 61495: contig of 4107 bp in length
* 61496 61595: gap of 100 bp
* 61596 65649: contig of 4054 bp in length
* 65650 65749: gap of 100 bp
* 65750 68290: contig of 2541 bp in length
* 68291 68390: gap of 100 bp
* 68391 72533: contig of 4143 bp in length
* 72534 72633: gap of 100 bp
* 72634 76438: contig of 3805 bp in length
* 76439 76538: gap of 100 bp
* 76539 81249: contig of 4711 bp in length
* 81250 81349: gap of 100 bp
* 81350 84922: contig of 3573 bp in length

```

```

* 84923 85022: gap of 100 bp
* 85023 89415: contig of 4393 bp in length
* 89416 89515: gap of 100 bp
* 89516 94613: contig of 5098 bp in length
* 94614 94713: gap of 100 bp
* 94714 99626: contig of 4913 bp in length
* 99627 99726: gap of 100 bp
* 99727 105173: contig of 5447 bp in length
* 105174 105273: gap of 100 bp
* 105274 110494: contig of 5221 bp in length
* 110495 110594: gap of 100 bp
* 110595 116499: contig of 5905 bp in length
* 116500 121923: contig of 5324 bp in length
* 121924 122023: gap of 100 bp
* 122024 128945: contig of 6922 bp in length
* 128946 129045: gap of 100 bp
* 129046 135023: contig of 5978 bp in length
* 135024 135123: gap of 100 bp
* 135124 141699: contig of 6576 bp in length
* 141700 141799: gap of 100 bp
* 141800 148947: contig of 7148 bp in length
* 148948 149047: gap of 100 bp
* 149048 161005: contig of 11958 bp in length
* 161006 173539: contig of 12434 bp in length
* 173540 173639: gap of 100 bp
* 173640 183616: contig of 9977 bp in length
* 183617 183716: gap of 100 bp
* 183717 194718: contig of 11002 bp in length.

```

## FEATURES

## Location/Qualifiers

```

1..194718
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone_lib="RP11-768B23"
1..1164
/feature="assembly_fragment"
1265..2302
/feature="assembly_fragment"
2403..3517
/feature="assembly_fragment"
3618..4687
/feature="assembly_fragment"
4788..6355
/feature="assembly_fragment"
6456..7718
/feature="assembly_fragment"
7819..9644
/feature="assembly_fragment"
9745..11368
/feature="assembly_fragment"
11469..13195
/feature="assembly_fragment"
13296..14769
/feature="assembly_fragment"

```

Query Match 76.8%; Score 19.2; DB 70; Length 194718;  
Best Local Similarity 87.5%; Pred. No. 73;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgtcgaagtcacagaattatt 25  
||||| ||| ||||| |||||

DB 95673 TTGTCACAAATGACAGAATTATT 95696

## RESULT 30

AC020772

LOCUS

DEFINITION

AC020772 198759 bp DNA

Homo sapiens clone RP11-710K17, WORKING DRAFT SEQUENCE, 29

unordered pieces.

HTG

05-APR-2000

```

AC020772
VERSION      AC020772.3  GI:7417754
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 198759)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-710K17
Unpublished
2 (bases 1 to 198759)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczyk,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 5, 2000 this sequence version replaced gi:6850459.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4348
Center clone name: 710_K_17
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 181544 bases at least Q40
Consensus quality: 189586 bases at least Q30
Consensus quality: 193248 bases at least Q20
Insert size: 194000; agarose-fp
Insert coverage: 195959; sum-of-contigs
Quality coverage: 4.0 in Q20 bases; agarose-fp
Quality coverage: 3.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1036: contig of 1036 bp in length
* 1037 1136: gap of 100 bp
* 1137 2371: contig of 1235 bp in length
* 2372 2471: gap of 100 bp
* 2472 3036: contig of 565 bp in length
* 3037 3136: gap of 100 bp
* 3137 5321: contig of 2185 bp in length
* 5322 5421: gap of 100 bp
* 5422 7352: contig of 1931 bp in length

```

---

```

7353 7452: gap of 100 bp
7453 9745: contig of 2293 bp in length
9746 9845: gap of 100 bp
9846 11851: contig of 2006 bp in length
11852 11951: gap of 100 bp
11952 13770: contig of 1819 bp in length
13771 13870: gap of 100 bp
13871 17767: contig of 3897 bp in length
17768 17867: gap of 100 bp
17868 21613: contig of 3746 bp in length
21614 21713: gap of 100 bp
21714 25648: contig of 3935 bp in length
25649 25748: gap of 100 bp
25749 30143: contig of 4395 bp in length
30144 30243: gap of 100 bp
30244 35539: contig of 5296 bp in length
35540 35639: gap of 100 bp
35640 39963: contig of 4324 bp in length
39964 40063: gap of 100 bp
40064 45751: contig of 5688 bp in length
45752 45851: gap of 100 bp
45852 52881: contig of 7030 bp in length
52882 52981: gap of 100 bp
52982 61044: contig of 8063 bp in length
61045 61144: gap of 100 bp
61145 69139: contig of 7995 bp in length
69140 69239: gap of 100 bp
69240 76404: contig of 7165 bp in length
76405 76504: gap of 100 bp
76505 83553: contig of 7049 bp in length
83554 83653: gap of 100 bp
83654 91363: contig of 7710 bp in length
91364 91463: gap of 100 bp
91464 100416: contig of 8953 bp in length
100417 100516: gap of 100 bp
100517 108752: contig of 8236 bp in length
108753 108852: gap of 100 bp
108853 118890: contig of 10038 bp in length
118891 118990: gap of 100 bp
118991 131762: contig of 12772 bp in length
131763 131862: gap of 100 bp
131863 143447: contig of 11585 bp in length
143448 143547: gap of 100 bp
143548 160191: contig of 16644 bp in length
160192 160291: gap of 100 bp
160292 180075: contig of 19784 bp in length
180076 180175: gap of 100 bp
180176 198759: contig of 18584 bp in length.
Location/Qualifiers
1. 198759
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-710K17"
/clone_lib="RPC1-11 Human Male BAC"
1. 1036
/note="assembly_fragment"
1137. 2371
/note="assembly_fragment"
2472. 3036
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
3137. 5321
/note="assembly_fragment"
5422. 7352
/note="assembly_fragment"
7453. 9745
/note="assembly_fragment"
9846. 11851
/note="assembly_fragment"
11952. 13770
/note="assembly_fragment"
13871. 17767

```



```

repeat_region      /rpt_family="L1MA4A"
15843. .15944
repeat_region      /rpt_family="Alusx"
15945. .15969
repeat_region      /rpt_family="(CAAAA)n"
complement(16603. .16676)
/rpt_family="(TAAA)n"
17273. .17299
repeat_region      /rpt_family="AT_rich"
complement(18248. .18399)
/rpt_family="MIR"
19474. .19532
repeat_region      /rpt_family="L1MA5"
complement(19740. .19815)
/rpt_family="MER20"
complement(19876. .20365)
/rpt_family="L2"
21517. .21886
repeat_region      /rpt_family="LTR16C"
22416. .22529
/rpt_family="L2"
24462. .24762
repeat_region      /rpt_family="Alusg"
25169. .25201
/rpt_family="AT_rich"
complement(25230. .25495)
/rpt_family="LTR16C"
26138. .26169
repeat_region      /rpt_family="POLY_A"
complement(26308. .26716)
/rpt_family="MSTA"
complement(27493. .27516)
/rpt_family="(TAAA)n"
complement(27517. .27816)
/rpt_family="Alusx"
complement(30094. .30125)
/rpt_family="(CA)n"
30776. .30875
repeat_region      /rpt_family="MIR"
complement(31703. .31768)
/rpt_family="(CATA)n"
complement(31892. .32060)
/rpt_family="MER58B"
complement(32351. .32597)
/rpt_family="MIR"
32697. .32765
repeat_region      /rpt_family="L2"
complement(33799. .33856)
/rpt_family="AT_rich"
34422. .34460
repeat_region      /rpt_family="AT_rich"
34912. .35021
/rpt_family="MIR"
35194. .35256
repeat_region      /rpt_family="(CA)n"
35293. .35749
/rpt_family="MLTIC"
36048. .36592
repeat_region      /rpt_family="TIGGER2"
36594. .36883
/rpt_family="Alu"
36884. .37105
repeat_region      /rpt_family="TIGGER2"
37160. .37408
/rpt_family="TIGGER2"
complement(37657. .38294)
/rpt_family="L1M4"
38710. .38755
repeat_region      /rpt_family="L1M4"
38755. .39465
/rpt_family="L1M6"
complement(39603. .40801)
/rpt_family="L1M4"

40801. .40852
/rpt_family="L1"
41310. .41864
repeat_region      /rpt_family="THE1B"
42024. .42058
/rpt_family="AT_rich"
43394. .43535
/rpt_family="MIR"
complement(43827. .43964)
/rpt_family="MIR"
44283. .44581
repeat_region      /rpt_family="L1PA7"
complement(44665. .44997)
/rpt_family="L1ME1"
complement(44901. .45198)
/note=Region: T02862 F013A8 Fetal brain, Stratagene Homo
sapiens cDNA clone F013A8 3'end"
complement(45177. .45247)
/rpt_family="MER5A"
45492. .45527
/rpt_family="AT_rich"
45645. .46308
/rpt_family="L1PA2"
46311. .46350
/rpt_family="AT_rich"
46443. .46660
/rpt_family="L1"
46658. .46731
/rpt_family="L1PA4"
47822. .48274
/rpt_family="L1PA15"
49737. .49813
/rpt_family="(TAA)n"
49856. .49910
/rpt_family="(CA)n"
complement(50429. .50548)
/rpt_family="L2"
51514. .51543
/rpt_family="AT_rich"

Query Match      76.8%; Score 19.2; DB 85; Length 207957;
Best Local Similarity 87.5%; Pred. No. 74;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtacagaattattt 25
||||| ||| ||||| ||||| |||
Db 85514 TTGTCACAAATCAGAGATTTT 85491

RESULT 32
AC007553/c
LOCUS      AC007553 225635 bp DNA HTG 23-FEB-2001
DEFINITION Homo sapiens chromosome 12 clone RP11-557K11, WORKING DRAFT
SEQUENCE, 20 unordered pieces.
ACCESSION  AC007553
VERSION     AC007553.26 GI:13096008
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 225635)
AUTHORS     Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

```

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsli,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lozsed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., and Gibbs,R.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Direct Submission

Unpublished

2 (bases 1 to 225635)

Worley, K.C.

Direct Submission

Submitted (14-MAR-1999) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Feb 22, 2001 this sequence version replaced gi:12830037.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: HMEM

Center clone name: RP11-557K11

----- Summary Statistics

Sequencing vector: M13; L08821

Chemistry: Dye-primer; Body: 62% of reads

Chemistry: Dye-terminator; Big Dye; 38% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 208689 bases at least Q40

Consensus quality: 215000 bases at least Q30

Consensus quality: 218088 bases at least Q20

Estimated insert size: 215078; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 6.2x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently  
consists of 20 contigs. The true order of the pieces

is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

as soon as it is available and the accession number will

\* be preserved.

\* 1 50936: contig of 50936 bp in length

\* 50937 51036: gap of unknown length

\* 51037 93576: contig of 42540 bp in length

\* 93577 93676: gap of unknown length

\* 93677 121032: contig of 27356 bp in length

\* 121033 121132: gap of unknown length

\* 121133 142037: contig of 20905 bp in length

\* 142038 142137: gap of unknown length

\* 142138 162042: contig of 19905 bp in length

\* 162043 162143: gap of unknown length

\* 162143 176395: contig of 14253 bp in length

\* 176395 176495: gap of unknown length

\* 176496 186423: contig of 9928 bp in length

\* 186424 186523: gap of unknown length

\* 186524 200263: contig of 13740 bp in length

\* 200264 200363: gap of unknown length

\* 200364 209341: contig of 8978 bp in length

\* 209342 211667: contig of 2226 bp in length

\* 211668 211768: gap of unknown length

\* 211768 213250: contig of 1483 bp in length

\* 213251 213350: gap of unknown length

\* 213351 214801: contig of 1451 bp in length

\* 214802 214901: gap of unknown length

\* 214902 215980: contig of 1079 bp in length

\* 215981 216080: gap of unknown length

\* 216081 217525: contig of 1445 bp in length

\* 217526 217625: gap of unknown length

\* 217626 218738: contig of 1113 bp in length

\* 218739 218838: gap of unknown length

\* 218839 220220: contig of 1382 bp in length

\* 220221 220320: gap of unknown length

\* 220321 221697: contig of 1377 bp in length

\* 221698 221797: gap of unknown length

\* 221798 223032: contig of 1235 bp in length

\* 223033 223132: gap of unknown length

\* 223133 224265: contig of 1133 bp in length

\* 224266 224365: gap of unknown length

\* 224366 225635: contig of 1270 bp in length.

## FEATURES

## source

1..225635  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP11-557K11"

BASE COUNT 68274 a 42559 c 43779 g 69089 t 1934 others  
ORIGIN

Query Match 76.8%; Score 19.2; DB 60; Length 225635;  
Best Local Similarity 87.5%; Pred. No. 74;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtcacagaattatt 24

||||||||||||||||

Db 88869 ATTTAGCAAGTCATAGAAATTAT 88846

RESULT 33

ALI33462

LOCUS

DEFINITION

Human DNA sequence from clone RP11-108A16 on chromosome 20.

Contains STSS and GSSs, complete sequence.

ACCESSION ALI33462

VERSION ALI33462.23

KEYWORDS

HTG.

SOURCE

ORGANISM

human.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (14-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
On Aug 21, 2000 this sequence version replaced gi:9801283.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/Projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr20>

IMPORTANT: This sequence is not the entire insert of clone RP11-108A16 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP5-106901 is at 84379 in this sequence. The true right end of clone RP4-790B6 is at 100 in this sequence.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP11-108A16 is from the library RP11-11.1 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6.

#### FEATURES

##### source

Location/Qualifiers

1. .84478

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="20"

/clone="RP11-108A16"

/clone\_lib="RPC1-11.1"

441. .902

/note="MLT1E repeat: matches 54. .568 of consensus"

complement(863. .1274)

/note="match: GSS: Em:AQ409299"

complement(974. .1162)

/note="match: GSS: Em:AQ413763"

1293. .1648

/note="match: GSS: Em:AQ491347"

1359. .1672

/note="match: GSS: Em:AQ089882"

1405. .1705

/note="Aluub repeat: matches 1. .305 of consensus"

2354. .2542

/note="Tigter2a repeat: matches 1. .196 of consensus"

2543. .2833

/note="Alusp repeat: matches 1. .288 of consensus"

2834. .3094

/note="Tigter2a repeat: matches 196. .434 of consensus"

complement(4171. .4768)

/note="match: GSS: Em:AQ418133"

complement(4337. .4768)

/note="match: GSS: Em:AQ237510"

4363. .4672

/note="Alusc repeat: matches 1. .305 of consensus"

complement(4575. .4753)

/note="match: GSS: Em:AQ419566"

6427. .6488

/note="31 copies 2 mer tg 71% conserved"

7412. .7723

/note="Alusq repeat: matches 1. .311 of consensus"

9182. .9303

/note="MER3 repeat: matches 75. .209 of consensus"

9542. .9863

/note="MER33 repeat: matches -9. .321 of consensus"

11769. .11855

/note="MER97b repeat: matches 948. .1037 of consensus"

11871. .12581

/note="LIMB6 repeat: matches 5446. .6164 of consensus"

misc\_feature

misc\_feature

repeat\_region

repeat\_region

misc\_feature

misc\_feature

misc\_feature

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

12143. .12710

/note="match: GSS: Em:AQ308637"

complement(12249. .12756)

/note="match: GSS: Em:AQ118115"

12582. .13258

/note="MER97c repeat: matches 10. .777 of consensus"

13541. .14065

/note="MLT1F repeat: matches 13. .541 of consensus"

complement(15380. .15816)

/note="match: GSS: Em:AQ031863"

complement(15526. .15839)

/note="match: STS: Em:HSPF15A1"

16822. .16939

/note="Single clone region. Assembly consistent with restriction digest."

17530. .17567

/note="WADE1 repeat: matches 1. .38 of consensus"

18129. .18871

/note="LIM4 repeat: matches 2200. .2928 of consensus"

18937. .19474

/note="LIM4 repeat: matches 4674. .5232 of consensus"

20088. .20386

/note="Aluy repeat: matches 1. .303 of consensus"

20424. .20500

/note="Alu repeat: matches 1. .77 of consensus"

20501. .23893

/note="LIPAZ2 repeat: matches 2740. .6144 of consensus"

25257. .25937

/note="match: GSS: Em:A2305856"

27252. .27589

/note="match: GSS: Em:A2318983"

28250. .28413

/note="L2 repeat: matches 2538. .2705 of consensus"

28705. .29775

/note="LIPAZ2 repeat: matches 5073. .6144 of consensus"

30032. .30250

/note="MER30 repeat: matches 2. .223 of consensus"

30334. .30433

/note="Aluub repeat: matches 1. .104 of consensus"

complement(32162. .32669)

/note="match: STS: Em:G58112"

match: GSS: Em:AQ385474"

33084. .33309

/note="match: GSS: Em:AQ595367 Em:AQ669389"

33877. .34227

/note="LTR7 repeat: matches 1. .448 of consensus"

34228. .34607

/note="HERVH repeat: matches 7324. .7713 of consensus"

34608. .34844

/note="HERVH repeat: matches 5588. .5819 of consensus"

34825. .35487

/note="HERVH repeat: matches 4488. .5149 of consensus"

35490. .36344

/note="HERVH repeat: matches 3140. .3993 of consensus"

36324. .39331

/note="HERVH repeat: matches 1. .2995 of consensus"

37253. .37773

/note="Sequence from clone PCR. Assembly is consistent with restriction digest."

complement(37712. .37773)

/note="Weak data"

37791. .37829

/note="Single clone region. Assembly consistent with restriction digest."

39334. .39684

/note="LTR7 repeat: matches 1. .448 of consensus"

39779. .40059

/note="LIME3A repeat: matches 5836. .6115 of consensus"

40657. .41016

/note="MLT1A1 repeat: matches 1. .365 of consensus"

41114. .41167

/note="MIR repeat: matches 13. .66 of consensus"

41504. .41906



```

repeat_region /note="L2 repeat: matches 2248. .2667 of consensus"
43254. 43645
repeat_region /note="MLT1A2 repeat: matches 4. .374 of consensus"
43865. .44012
repeat_region /note="L1R33 repeat: matches 76. .256 of consensus"
44137. .44198
repeat_region /note="L1R33 repeat: matches 431. .493 of consensus"
44537. 44698
repeat_region /note="MIR repeat: matches 53. .212 of consensus"
45920. .45967
repeat_region /note="24 copies 2 mer ac 79% conserved"
47253. .47427
repeat_region /note="MIR repeat: matches 53. .248 of consensus"
48266. 48374
repeat_region /note="MER86 repeat: matches 26. .143 of consensus"
48733. .49291
misc_feature /note="match: GSS: Em:AQ386870
match: STS: Em:G58335"
48852. .49191
misc_feature /note="match: STS: Em:G63817"
48856. .48978
repeat_region /note="L1MFC repeat: matches 1804. .1927 of consensus"
49342. .49517
repeat_region /note="L1MFC repeat: matches 2418. .2237 of consensus"
49798. .50003
repeat_region /note="MER58C repeat: matches 1. .89 of consensus"
50045. .50241
repeat_region /note="L2 repeat: matches 2553. .2744 of consensus"
50293. .50614
repeat_region /note="MER2 repeat: matches 1. .345 of consensus"
50615. .50795
repeat_region /note="L1M4 repeat: matches 2881. .3050 of consensus"
50796. .51092
repeat_region /note="AluJo repeat: matches 1. .294 of consensus"

Query Match 75.2%; Score 18.8; DB 89; Length 84478;
Best Local Similarity 90.9%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattat 23
||||| ||||||| |||||||
Db 64614 TTGTCCAAAGTCACAGAATTAT 64635

RESULT 34
AC012380 108623 bp DNA PRI 26-OCT-1999
LOCUS Genomic Sequence For Homo sapiens Clone 125H5, Chromosome 20,
DEFINITION complete sequence.
ACCESSION AC012380
VERSION AC012380.1 GI:6119509
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 108623)
AUTHORS Spigel, L.A., Nascimento, L.U., de la Bastide, M., Vil, D.M.,
Huang, E.N., Preston, R.R., Matero, A., Shah, R., O'Shaughnessy, A.,
Rodriguez, M., Shekher, M., Schutz, K., See, L.H., Swaby, I.,
Habermann, K., Dedhia, N.N. and McCombie, W.R.
TITLE Genomic Sequence For Homo sapiens Clone 125H5, Chromosome 20,
Complete Sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 108623)
AUTHORS McCombie, W.R.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
FEATURES
source 1. 108623
/organism="Homo sapiens"

```

```

/db_xref="taxon:9606"
/chromosome="20"
/clone="125H5"
56550. .56820
/note="The assembly is single stranded with single
chemistry."
BASE COUNT 33048 a 19822 c 21292 g 34461 t
ORIGIN

Query Match 75.2%; Score 18.8; DB 87; Length 108623;
Best Local Similarity 90.9%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattat 23
||||| ||||||| |||||||
Db 39442 TTGTCCAAAGTCACAGAATTAT 39463

RESULT 35
AC064870/c 193443 bp DNA HTG 01-SEP-2000
LOCUS Homo sapiens chromosome 2 clone RP11-575E20, WORKING DRAFT
DEFINITION SEQUENCE, 22 unordered pieces.
ACCESSION AC064870
VERSION AC064870.2 GI:9958289
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 193443)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 193443)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (23-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Sep 1, 2000 this sequence version replaced gi:7637824.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0575E20
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 97% of reads
Chemistry: Dye-terminator Big Dye; 3% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 186954 bases at least Q40
Consensus quality: 188592 bases at least Q30
Consensus quality: 189449 bases at least Q20
Insert size: 202000; agarose-fp
Insert size: 191343; sum-of-contigs
Quality coverage: 4.14 in Q20 bases; agarose-fp
Quality coverage: 4.40 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 3288: contig of 3288 bp in length
* 3289 3388: gap of unknown length

```

```

* 3389: contig of 4002 bp in length
* 7391: gap of unknown length
* 7491: contig of 5268 bp in length
* 12758: gap of unknown length
* 12859: contig of 5263 bp in length
* 18121: gap of unknown length
* 18221: contig of 6088 bp in length
* 24309: gap of unknown length
* 24409: contig of 5737 bp in length
* 30146: gap of unknown length
* 30246: contig of 9217 bp in length
* 39463: gap of unknown length
* 39563: contig of 9702 bp in length
* 49265: gap of unknown length
* 49365: contig of 9742 bp in length
* 49366: gap of unknown length
* 59107: contig of 9882 bp in length
* 59208: gap of unknown length
* 59209: contig of 12355 bp in length
* 69190: gap of unknown length
* 81544: contig of 10789 bp in length
* 81545: gap of unknown length
* 92433: contig of 13828 bp in length
* 92434: gap of unknown length
* 106361: contig of 15517 bp in length
* 106461: gap of unknown length
* 121978: contig of 15517 bp in length
* 122078: gap of unknown length
* 141403: contig of 19325 bp in length
* 141503: gap of unknown length
* 141504: contig of 17361 bp in length
* 158864: gap of unknown length
* 158964: gap of unknown length
* 183594: contig of 24630 bp in length
* 183694: gap of unknown length
* 185169: contig of 1475 bp in length
* 185269: gap of unknown length
* 185270: contig of 1481 bp in length
* 186750: gap of unknown length
* 186751: contig of 1921 bp in length
* 188771: gap of unknown length
* 188772: contig of 2270 bp in length
* 191141: gap of unknown length
* 191241: contig of 2202 bp in length.
* 191242: 193443: contig of 2202 bp in length.

```

## FEATURES

source

```

1. .193443
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-575E20"
1. .3288
/note="assembly_name:Contig10"
3389. .7390
/note="assembly_name:Contig11"
7491. .12758
/note="assembly_name:Contig12"
12859. .18121
/note="assembly_name:Contig13"
18222. .24309
/note="assembly_name:Contig14"
24410. .30146
/note="assembly_name:Contig15"
30247. .39463
/note="assembly_name:Contig16"
39564. .49265
/note="assembly_name:Contig17"
49366. .59107
/note="assembly_name:Contig18"
59208. .69089
/note="assembly_name:Contig19"
69190. .81544
/note="assembly_name:Contig20"
81645. .92433
/note="assembly_name:Contig21"
92534. .106361
/note="assembly_name:Contig22"

```

```

misc_feature 106462. .121978
/note="assembly_name:Contig23"
clone_end:w7
vector_side:right"
misc_feature 122079. .141403
/note="assembly_name:Contig24"
misc_feature 141504. .158864
/note="assembly_name:Contig25"
misc_feature 158965. .183594
/note="assembly_name:Contig26"
misc_feature 183695. .185169
/note="assembly_name:Contig5"
misc_feature 185270. .186750
/note="assembly_name:Contig6"
misc_feature 186851. .188771
/note="assembly_name:Contig7"
misc_feature 188872. .191141
/note="assembly_name:Contig8"
clone_end:SP6
vector_side:left"
misc_feature 191242. .193443
/note="assembly_name:Contig9"

```

```

BASE COUNT 60070 a 35783 c 36301 g 59178 t 2111 others
ORIGIN

```

```

Query Match 75.2%; Score 18.8; DB 72; Length 193443;
Best Local Similarity 90.9%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 4 gtgcgaagtcacagaattattt 25
|| ||||| ||||| |||
Db 47787 GTTCAAGTCACAGAATTGTTT 47766

```

## RESULT 36

```

LOCUS A75959 2114 bp DNA PAT 19-OCT-1999
DEFINITION Sequence 1 from Patent WO9321326.
ACCESSION A75959
VERSION A75959.1 GI:6088149
KEYWORDS
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 2114)
Kazmaier, M. and Lacroute, F.
METHOD OF CLONING A DNA SEQUENCE CODING FOR A PLANT
NADPH-CYTOCHROME P450 REDUCTASE AND ASSOCIATED DNA SEQUENCES
Patent: WO 9321326-A 1 28-OCT-1993;
ORGAN (FR); KAZMAIER MICHAEL (FR)

```

## FEATURES

```

source
Location/Qualifiers
1. .2114
/organism="Arabidopsis thaliana"
/strain="LANSBERG ERECTA"
/db_xref="taxon:3702"
/dev_stage="JEUNE PLANTULE STADE DEUX"
/dev_stage="FEUILLES"
1. .35
5'UTR
gene
36. .2114
/clone="ARA B"
CDS
36. .2114
/clone="ARA B"
/codon_start=1
/product="NADPH CYTOCHROME P450 REDUCTASE"
/protein_id="CAB98575.1"
/db_xref="GI:6088150"
/translation="MTSALYASDLFKQLKSLMGTDLSDDVLIATTSALYAGFVY
LLWKTTADRSGLPLMKSLMKDDEDDLDLGSKTRVSIFFGTGTGTAEGFARA
LSEETKAYEKAAYKVIDDDYAADDDQYEKLAKETLAFPCVATYGDGPTDNAARF
SKWFTENERDINKLQQLAYGAYFALGNROYEHFNKIGIVLDELCKKAKRLIEVLGD

```

DDOSIEDFNWAKESLWSELDKLLKDDKSVATPYTAVIPEVRVYTHDPRTTOKSM  
 ESNVANGNTIDIHPCRVDAVQKELHSHEDRSCHLHLEFEDISRTGTYEYEGDHWGV  
 YANHEIVEAEKLGHLSDLVFSIADKEDGSPLESAPVPPFPCTGLGTGLARIA  
 DLLNPRKSAVALAAYTEPSEAKLHLTSPDGKDEYSQWIVASQSLLEVMAAF  
 SAKPPLGLVFAAPLQPIRYISISSCDWAPSRVHTVSVLVYGTPTGRIHKGVCST  
 WMKNVAPAEKSHSCGAPIFIRASNEKLPSPSTPIVMVPGTGLAPFGRFLOERMAL  
 KEDGEELGSSLLFPFCGRNOMDFIYEDELNMFVDOGVISELIMAFSRECAQKEYVOHK  
 MMEKAAOVMDLIKEBGYLYCGDAKGMARDVHRTLHTIVQEQGVSSSSEAEIVKKLQ  
 TEGRYLRDWM"  
 2111..2114  
 /gene="ARA B"  
 BASE COUNT 603 a 399 c 542 g 570 t  
 ORIGIN

Query Match 74.4%; Score 18.6; DB 9; Length 2114;  
 Best Local Similarity 84.0%; Pred. No. 1.3e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtcacagaattattt 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 1338 ATTGTTGCAAGTCAGAGAAGTCTTT 1362

RESULT 37  
 ATATR1G  
 LOCUS ATATR1G 2199 bp mRNA PLN 01-JUN-1992  
 DEFINITION A.thaliana ATR1 mRNA for NADPH-cytochrome P450 reductase.  
 ACCESSION X66016  
 VERSION X66016.1 GI:16186  
 KEYWORDS ATR1 gene; NADPH-ferrihemoprotein reductase.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 Brassicales; Brassicaceae; Arabidopsids.  
 1 (bases 1 to 2199)  
 Pompon,D.M.  
 Direct Submission  
 Submitted (11-MAY-1992) D.M. Pompon, Centre de Genetique Mol du  
 CNRS, Avenue de la Terrasse, 91198 Gif-Sur-Yvette Cedex, FRANCE  
 2 (bases 1 to 2199)  
 Mignote-Vieux,C., Kazmaier,M., Lacroute,F. and Pompon,D.M.  
 Unpublished  
 Location/Qualifiers  
 1..2199  
 /organism="Arabidopsis thaliana"  
 /strain="Heynh"  
 /sub\_strain="Landsberg erecta"  
 /db\_xref="taxon:3702"  
 /dev\_stage="seedling (2 leaves)"  
 /tissue\_type="whole seedling"  
 70..2148  
 /gene="ATR1"  
 70..2148  
 /gene="ATR1"  
 /note="NADPH-cytochrome P450 reductase"  
 /codon\_start=1  
 /product="NADPH-ferrihemoprotein reductase"  
 /protein\_id="CAA6814.1"  
 /db\_xref="GI:16187"  
 /db\_xref="SPTREMBL:Q39035"  
 /translation="MTSALYASDLFLKOLKSIIMGTDLSDDVLVLIATTSALVAGFVV  
 LLMKTTADRSGELKLMIPKSLMAKDEDDLDLGSCTRVSIFFGTGTAGFAKA  
 LSEETIKARYEAAVKVITDLDYAADDOVEEKLKTELAPFCVATYGDGTEPTNAARF  
 SKWETENEDILKQOLAYGVAFNQEYEFNKIGIVLDELCKKAGKLEVLGLD  
 DDQSIDFNWAKESLWSELDKLLKDDKSVATPYTAVIPEVRVYTHDPRTTQSM  
 ESNVANGNTIDIHPCRVDAVQKELHSHEDRSCHLHLEFEDISRTGTYEYEGDHWGV  
 YANHEIVEAEKLGHLSDLVFSIADKEDGSPLESAPVPPFPCTGLGTGLARIA  
 DLLNPRKSAVALAAYTEPSEAKLHLTSPDGKDEYSQWIVASQSLLEVMAAF  
 SAKPPLGLVFAAPLQPIRYISISSCDWAPSRVHTVSVLVYGTPTGRIHKGVCST  
 WMKNVAPAEKSHSCGAPIFIRASNEKLPSPSTPIVMVPGTGLAPFGRFLOERMAL  
 KEDGEELGSSLLFPFCGRNOMDFIYEDELNMFVDOGVISELIMAFSRECAQKEYVOHK

gene  
 CDS

MMEKAAOVMDLIKEBGYLYCGDAKGMARDVHRTLHTIVQEQGVSSSSEAEIVKKLQ  
 TEGRYLRDWM"  
 BASE COUNT 619 a 420 c 552 g 608 t  
 ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 2199;  
 Best Local Similarity 84.0%; Pred. No. 1.3e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtcacagaattattt 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 1372 ATTGTTGCAAGTCAGAGAAGTCTTT 1396

RESULT 38  
 CEF19C6  
 LOCUS CEF19C6 36493 bp DNA INV 25-OCT-2000  
 DEFINITION Caenorhabditis elegans cosmid F19C6, complete sequence.  
 ACCESSION Z48006  
 VERSION Z48006.1 GI:642183  
 KEYWORDS HTG; EGF-like domain; G protein-coupled receptor kinase.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 1 (bases 1 to 36493)  
 none.  
 HARRIS,B.R.  
 Direct Submission  
 Submitted (28-JAN-1995) Nematode Sequencing Project, Sanger Centre,  
 Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,  
 Washington University, St. Louis, MO 63110, USA. E-mail:  
 jes@sanger.ac.uk or rw@nematode.wustl.edu  
 Coding sequences below are predicted from computer analysis, using  
 predictions from GeneFinder (P. Green, U. Washington), and other  
 available information.  
 Current sequence finishing criteria for the C. elegans genome  
 sequencing consortium are that all bases are either sequenced  
 unambiguously on both strands, or on a single strand with both a  
 dye primer and dye terminator reaction, from distinct subclones.  
 Exceptions are indicated by an explicit note.  
 IMPORTANT: This sequence is not the entire insert of clone F19C6.  
 It may be shorter because we only sequence overlapping sections  
 once, or longer because we arrange for a small overlap between  
 neighbouring submissions.  
 The true left end of clone F19C6 is at 1 in this sequence. The true  
 right end of clone F19C6 is at 104 in  
 sequence Z48230.  
 The true left end of clone F42G10 is at 9861 in this sequence. The  
 start of this sequence (1..101) overlaps with the end of sequence  
 Z47073.  
 The end of this sequence (36390..36493) overlaps with the start of  
 sequence Z48230.  
 For a graphical representation of this sequence and its analysis  
 see: [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F19C6)  
 name=F19C6.  
 Location/Qualifiers  
 1..36493  
 /organism="Caenorhabditis elegans"  
 /db\_xref="taxon:6239"  
 /chromosome="X"  
 /clone="F19C6"  
 complement(join(2639..2779,3185..3315,3994..4072,

FEATURES  
 Source  
 gene

```

1116. 4334,4480. .4537,4590. .4673,4915. .4958))
/ gene="F19C6.5"
complement(join(2639. .2779,3185. .3315,3994. .4072,
4116. 4334,4480. .4537,4590. .4673,4915. .4958))
/ gene="F19C6.5"
/codon_start=1
/protein_id="CAA88049.1"
/db_xref="GI:3876081"
/db_xref="SWISS-PROT:Q09308"
/translation="MTEKDEYGTIVVARVAMEENFENPQDIPRIKNDSTHSYQOPE
SMLEKNTVYTLAVVLAQTLIGVLSNLTIVADATRLADHLELFQYDRAATPGKRLT
EIITQNTVILFVAFLLTASGRATWFEINRLISIGTAMAMTANTLIQICHF
RTWQREYSSHTSYKQNLMDHGFYHFIAYFLVSSLLIIVNKDYLIADVITTYA
TSLLIANSISQYLYHEYFSLEHPQDEYPI"
complement(join(12098. .12238,13715. .13867,14218. .14379,
14432. .14591,14835. .15039,15263. .15581,15692. .15817,
16720. .16891,17784. .17914,17967. .18068,18651. .18756,
18882. .18981,19084. .19135))
/ gene="F19C6.1"
complement(join(12098. .12238,13715. .13867,14218. .14379,
14432. .14591,14835. .15039,15263. .15581,15692. .15817,
16720. .16891,17784. .17914,17967. .18068,18651. .18756,
18882. .18981,19084. .19135))
/ note="contains similarity to Pfam domain: PF00069
(Eukaryotic protein kinase domain), Score=257.8,
E-value=4.7e-74, N=1; PF00615 (Regulator of G protein
signaling domain), Score=156.0, E-value=2.1e-43, N=1
cDNA EST yk21f11.3 comes from this gene
cDNA EST yk2a1.5 comes from this gene
cDNA EST yk21f11.5 comes from this gene
cDNA EST yk33f1.5 comes from this gene
cDNA EST yk246c6.5 comes from this gene
cDNA EST yk573a3.5 comes from this gene"
/codon_start=1
/protein_id="CAA88047.1"
/db_xref="GI:3876079"
/db_xref="SWISS-PROT:Q09537"
/translation="MEINIVANTVYIKARESOGKKGKKWKNYLPQPHYTECLPL
KTEIVTYAVRQKQIGLKLHFEQCOATNSHQCCFLTKVYEETEDDDQSRKD
LASVLSLHKSDEPNSQDEQFCWSFLSDEVISTCITADSATHDDEDSFT
EPYRLCOYLTDAPEKFAESMYFHRLQWKLKRPVDKHTFRLYRLGKGFGVE
ACQVRSKMYALKLEKFKRKHAEATLSLNEQLQIRNSPFSVLAAYETDAL
CLVTLMHGSDLHLYNLMPGDFEQRQVFAAEITLGLHLSRIYLRQKPENI
LLDDFHVGRDVLGLAVEIKDNEPTIKRGVTVGYMAPEIVKNERYSYGVWGVGCL
YEMLEKAPRQKREKVEVERRYRDEQEKSEKSEAAATLCRGLLHKEGPRLG
CRVGRPEDAEIRAHFPENTADTVTGREPVPWKMEAGKVTTPPCDPRAVYKDV
LDLEQSTVGVRLDADTDQFYGKFNCGVSPWQSEMIETECFAELNFFHDEGNYM
WNLRFQGINNDRNRNKGPGFSLRFRKKNIEVTKSLHDLHLGVEQQPPKTSQT
PAVRSRRAASAGRTLVI"
join(21023. .21152,21233. .21453,22601. .22750,22796. .22870,
23454. .23739,23798. .23957,24236. .24326,24747. .24983)
/ gene="F19C6.2b"
join(21023. .21152,21233. .21453,22601. .22750,22796. .22870,
23454. .23739,23798. .23957,24236. .24326,24747. .24983)
/ gene="F19C6.2b"
join(21023. .21152,21233. .21453,22601. .22750,22796. .22870,
23454. .23739,23798. .23957,24236. .24326,24747. .24983)
/ gene="F19C6.2a"
join(21023. .21152,21233. .21453,22601. .22750,22796. .22870,
23454. .23739,23798. .23957,24236. .24326,24747. .24983)
/ gene="F19C6.2b"
join(21023. .21152,21233. .21453,22601. .22750,22796. .22870,
23454. .23739,23798. .23957,24236. .24326,24747. .24983)
/ note="cDNA EST yk630f11.5 comes from this gene"
/codon_start=1
/protein_id="CAB62803.1"
/db_xref="GI:6562357"
/db_xref="SPTREMBL:Q90315"
/translation="MAEKILFYIYRDSCTSSSENLYRAIVDEKFNVFRDRGIATFSKG
YSGIPLGELPIDSDLDLWVAVLFIKISAREEGLIFKVHTTDEKFNPLVAEAKSD
VLPNWASHVKDFEAENFVGMNGVIHLISPTVSEKPKRSQIDILRAVNPENILVATES
LSEKILHSSSEKFNMGHNTSNVAVPYMAVERPDILSLAIFREYVDPDKIKLELR
LANEHRVMIHTIINDTDWKEVTAVADIESPTDIVSHRVSLALLAFDEKISSPENGVD
VPSGLFQKVGDRFEREHLALRAKLFQSPSATHLYOCAKSLVTGNHVOBCRKFVD
EGSTSNRSDCSGNEEDSSNAKYAKQVKKKKRKAFAKRELAAILDHTKPSHSEP
DYVAPDNEALKKFERAVNGDDYKASDEYSLGEEDLELFMAAPRKRIIEEMKNKN
KARLGIKVTERLPTVSDDEEDVADMLRAAPPIANAEDIEDFDDI"
complement(join(25960. .26079,26631. .26702,26761. .26932,
26976. .27082,27153. .27249,27297. .27367,27529. .27603,
27872. .28037,28080. .28276,28627. .28839))
/ gene="F19C6.3"
complement(join(25960. .26079,26631. .26702,26761. .26932,
26976. .27082,27153. .27249,27297. .27367,27529. .27603,
27872. .28037,28080. .28276,28627. .28839))
/ note="possesses an EGF-like domain, contains similarity
to Pfam domain: PF00008 (EGF-like domain), Score=33.7,
E-value=1.4e-06, N=2"
/codon_start=1
/protein_id="CAA88048.1"
/db_xref="GI:3876080"
/db_xref="SPTREMBL:Q09538"
/translation="MWAHLILPLVLSVSHCHIDRSYILAHVNNLLSTGESWILTN
FDLSNGTFFLAQNSERRIERGFKNTLYMVLPISKRIQSNWKEKRYVDFDLSIGKI
SGWLEITFNHNIHYVNCDEVISDFEHLQTLNHLPSLHIVNWKGPETQFLIG
RNPTQKQKHGDKVKKYADAYTEKIERNEPSAIDAINIEFFTVVQVPPFGDCKI
NNALINNETVFNWNTCTCVDRNVNQCQAVDCPAVCTHPMI RKDDCCSCGQCKY
ESHKIANPHGVFVWPCDQYRCQCKWKKVHCSSEYATCPCSPEDWYNEKFNCCPK
CRDFAFCVAPKCHYATCTDKSRKPCSDPTFGQNGTYCEDIDECFSQDAKEQLG
GCLSGTSCNVPVGSYKCDPLGPGMIGENTCLLIRV"
join(complement(248230.1:1308. .1367),
complement(248230.1:1111. .1227),
complement(248230.1:851. .1066),
complement(248230.1:268. .327), complement(248230.1:105),
complement(36284. .36493), complement(36093. .36239),
complement(35772. .35885), complement(35359. .35535),
complement(35174. .35317))
/ gene="F19C6.4"
join(complement(248230.1:1308. .1367),
complement(248230.1:1111. .1227),
complement(248230.1:851. .1066),
complement(248230.1:502. .803),
complement(248230.1:268. .327), complement(248230.1:105),
complement(36284. .36493), complement(36093. .36239),
complement(35772. .35885), complement(35359. .35535),
complement(35174. .35317))
/ note="contains similarity to pfam domain: PF01431
(Peptidase family M13), Score=268.8, E-value=2.4e-77, N=1"
/codon_start=1
/protein_id="CAA88051.1"
/db_xref="GI:3876083"
/db_xref="SPTREMBL:Q09539"
/translation="MKGFIFIVLTVTSQSFHKSIDISPCNDYDYVCAKDTGRI
SNLFTKMNQLEILKPNVTIENNGTFLADKIVKLRVMKWCETIEVLFTGVTVDY

```

Query Match 74.4%; Score 18.6; DB 6; Length 36493;

Best Local Similarity 84.0%; Pred. No. 1.4e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 attgtcgaagtcacagaattattt 25

DB 34465 AATGTTGCAAGTCAAAAAATTATTT 34489

# RESULT 39

AC023180/c

LOCUS AC023180 64218 bp DNA HTG 01-APR-2001

DEFINITION Homo sapiens clone RP11-16N4, LOW-PASS SEQUENCE SAMPLING.

ACCESSION AC023180

VERSION AC023180.3 GI:13493131

KEYWORDS HTG; HTGS\_PHASE0.

SOURCE human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Homo sapiens, clone RP11-16N4

JOURNAL Unpublished

## REFERENCE

AUTHORS

2316 2415: gap of 100 bp  
2416 3146: contig of 731 bp in length  
3147 3246: gap of 100 bp  
3247 3954: contig of 708 bp in length  
3955 4054: gap of 100 bp  
4055 4763: contig of 709 bp in length  
4764 4863: gap of 100 bp  
4864 5885: contig of 722 bp in length  
5886 5685: gap of 100 bp  
5686 6412: contig of 727 bp in length  
6413 6512: gap of 100 bp  
6513 7201: contig of 689 bp in length  
7202 7301: gap of 100 bp  
7302 8005: contig of 704 bp in length  
8006 8105: gap of 100 bp  
8106 8823: contig of 718 bp in length  
8824 8923: gap of 100 bp  
8924 9641: contig of 718 bp in length  
9642 9741: gap of 100 bp  
9742 10348: contig of 607 bp in length  
10349 10448: gap of 100 bp  
10449 11173: contig of 725 bp in length  
11174 11273: gap of 100 bp  
11274 11992: contig of 719 bp in length  
11993 12092: gap of 100 bp  
12093 12767: contig of 675 bp in length  
12768 12867: gap of 100 bp  
12868 13569: contig of 702 bp in length  
13570 13669: gap of 100 bp  
13670 14362: contig of 693 bp in length  
14363 14462: gap of 100 bp  
14463 15184: contig of 722 bp in length  
15185 15284: gap of 100 bp  
15285 16003: contig of 719 bp in length  
16004 16103: gap of 100 bp  
16104 16827: contig of 724 bp in length  
16828 16927: gap of 100 bp  
16928 17636: contig of 709 bp in length  
17637 17736: gap of 100 bp  
17737 18449: contig of 713 bp in length  
18450 18549: gap of 100 bp  
18550 19266: contig of 717 bp in length  
19267 19366: gap of 100 bp  
19367 20090: contig of 724 bp in length  
20091 20190: gap of 100 bp  
20191 20923: contig of 733 bp in length  
20924 21023: gap of 100 bp  
21024 21756: contig of 733 bp in length  
21757 21856: gap of 100 bp  
21857 22575: contig of 719 bp in length  
22576 22675: gap of 100 bp  
22676 23391: contig of 716 bp in length  
23392 23491: gap of 100 bp  
23492 24210: contig of 719 bp in length  
24211 24310: gap of 100 bp  
24311 25025: contig of 715 bp in length  
25026 25125: gap of 100 bp  
25126 25847: contig of 722 bp in length  
25848 25947: gap of 100 bp  
25948 26675: contig of 728 bp in length  
26676 26775: gap of 100 bp  
26776 27512: contig of 737 bp in length  
27513 27612: gap of 100 bp  
27613 28337: contig of 725 bp in length  
28338 28437: gap of 100 bp  
28438 29132: contig of 695 bp in length  
29133 29232: gap of 100 bp  
29233 29939: contig of 707 bp in length  
29940 30039: gap of 100 bp  
30040 30764: contig of 725 bp in length  
30765 30864: gap of 100 bp  
30865 31568: contig of 704 bp in length  
31569 31668: gap of 100 bp

## TITLE

JOURNAL

## COMMENT

Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 1, 2001 this sequence version replaced gi:7144956.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L3592

Center clone name: 16\_N\_4

-----

\* NOTE: This record contains 79 individual

\* sequencing reads that have not been assembled into

\* contigs. Runs of N are used to separate the reads

\* and the order in which they appear is completely

\* arbitrary. Low-pass sequence sampling is useful for

\* identifying clones that may be gene-rich and allows

\* overlap relationships among clones to be deduced.

\* However, it should not be assumed that this clone

\* will be sequenced to completion. In the event that

\* the record is updated, the accession number will

\* be preserved.

\* 1 598: contig of 698 bp in length

\* 699 798: gap of 100 bp

\* 799 1501: contig of 703 bp in length

\* 1502 1601: gap of 100 bp

\* 1602 2315: contig of 714 bp in length

```

* 31569 32387: contig of 719 bp in length
* 32388 32487: gap of 100 bp
* 32488 33201: contig of 714 bp in length
* 33202 33301: gap of 100 bp
* 33302 34014: contig of 713 bp in length
* 34015 34114: gap of 100 bp
* 34115 34821: contig of 707 bp in length
* 34822 34921: gap of 100 bp
* 34922 35649: contig of 728 bp in length
* 35650 35749: gap of 100 bp
* 35750 36485: contig of 736 bp in length
* 36486 36585: gap of 100 bp
* 36586 37313: contig of 728 bp in length
* 37314 37413: gap of 100 bp
* 37414 38133: contig of 720 bp in length
* 38134 38233: gap of 100 bp
* 38234 38956: contig of 723 bp in length
* 38957 39056: gap of 100 bp
* 39057 39763: contig of 707 bp in length
* 39764 39863: gap of 100 bp
* 39864 40571: contig of 708 bp in length
* 40572 40671: gap of 100 bp
* 40672 41383: contig of 712 bp in length
* 41384 41483: gap of 100 bp
* 41484 42205: contig of 722 bp in length
* 42206 42305: gap of 100 bp
* 42306 43034: contig of 729 bp in length
* 43035 43134: gap of 100 bp
* 43135 43853: contig of 719 bp in length
* 43854 43953: gap of 100 bp
* 43954 44692: contig of 739 bp in length
* 44693 44792: gap of 100 bp
* 44793 45519: contig of 727 bp in length
* 45520 45619: gap of 100 bp
* 45620 46314: contig of 695 bp in length
* 46315 46414: gap of 100 bp
* 46415 47117: contig of 703 bp in length
* 47118 47217: gap of 100 bp
* 47218 47943: contig of 726 bp in length
* 47944 48043: gap of 100 bp
* 48044 48745: contig of 702 bp in length
* 48746 48845: gap of 100 bp
* 48846 49559: contig of 714 bp in length
* 49560 49659: gap of 100 bp
* 49660 50385: contig of 726 bp in length
* 50386 50485: gap of 100 bp
* 50486 51208: contig of 723 bp in length
* 51209 51308: gap of 100 bp
* 51309 52007: contig of 699 bp in length
* 52008 52107: gap of 100 bp
* 52108 52779: contig of 672 bp in length
* 52780 52879: gap of 100 bp
* 52880 53609: contig of 730 bp in length
* 53610 53709: gap of 100 bp
* 53710 54451: contig of 742 bp in length
* 54452 54551: gap of 100 bp
* 54552 55282: contig of 731 bp in length
* 55283 55382: gap of 100 bp
* 55383 56094: contig of 712 bp in length
* 56095 56194: gap of 100 bp
* 56195 56886: contig of 692 bp in length
* 56887 56986: gap of 100 bp
* 56987 57710: contig of 724 bp in length
* 57711 57810: gap of 100 bp

```

Query Match 74.4%; Score 18.6; DB 67; Length 64218;  
 Best Local Similarity 84.0%; Pred. No. 1.4e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 attgtcgaagtcacagaattattt 25  
 ||||| ||||| ||||| ||||| |||||  
 DB 57601 ATTGTGCAATGACAGGATTATT 57577

RESULT 40  
 AC090618/c

LOCUS  
 DEFINITION

AC090618  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

AC090618.1  
 AC090618

```
* 3196 3890: contig of 695 bp in length
* 3891 3990: gap of 100 bp
* 3991 4700: contig of 710 bp in length
* 4701 4800: gap of 100 bp
* 4801 5523: contig of 723 bp in length
* 5524 5623: gap of 100 bp
* 5624 6344: contig of 721 bp in length
* 6345 6444: gap of 100 bp
* 6445 7159: contig of 715 bp in length
* 7160 7259: gap of 100 bp
* 7260 7985: contig of 726 bp in length
* 7986 8085: gap of 100 bp
* 8086 8794: contig of 709 bp in length
* 8795 8894: gap of 100 bp
* 8895 9604: contig of 710 bp in length
* 9605 9704: gap of 100 bp
* 9705 10415: contig of 711 bp in length
* 10416 10515: gap of 100 bp
* 10516 11270: contig of 755 bp in length
* 11271 11370: gap of 100 bp
* 11371 12111: contig of 741 bp in length
* 12112 12211: gap of 100 bp
* 12212 12918: contig of 707 bp in length
* 12919 13018: gap of 100 bp
* 13019 13724: contig of 706 bp in length
* 13725 13824: gap of 100 bp
* 13825 14535: contig of 711 bp in length
* 14536 14635: gap of 100 bp
* 14636 15356: contig of 721 bp in length
* 15357 15456: gap of 100 bp
* 15457 16170: contig of 714 bp in length
* 16171 16270: gap of 100 bp
* 16271 16968: contig of 698 bp in length
* 16969 17068: gap of 100 bp
* 17069 17790: contig of 722 bp in length
* 17791 17890: gap of 100 bp
* 17891 18599: contig of 709 bp in length
* 18600 18699: gap of 100 bp
* 18700 19388: contig of 689 bp in length
* 19389 19488: gap of 100 bp
* 19489 20208: contig of 720 bp in length
* 20209 20308: gap of 100 bp
* 20309 21048: contig of 740 bp in length
* 21049 21148: gap of 100 bp
* 21149 21847: contig of 699 bp in length
* 21848 21947: gap of 100 bp
* 21948 22662: contig of 715 bp in length
* 22663 22762: gap of 100 bp
* 22763 23479: contig of 717 bp in length
* 23480 23579: gap of 100 bp
* 23580 24288: contig of 709 bp in length
* 24289 24388: gap of 100 bp
* 24389 25117: contig of 729 bp in length
* 25118 25217: gap of 100 bp
* 25218 25922: contig of 705 bp in length
* 25923 26022: gap of 100 bp
* 26023 26765: contig of 743 bp in length
* 26766 26865: gap of 100 bp
* 26866 27601: contig of 736 bp in length
* 27602 27701: gap of 100 bp
* 27702 28413: contig of 712 bp in length
* 28414 28513: gap of 100 bp
* 28514 29246: contig of 733 bp in length
* 29247 29346: gap of 100 bp
* 29347 30085: contig of 739 bp in length
* 30086 30185: gap of 100 bp
* 30186 30897: contig of 712 bp in length
* 30898 30997: gap of 100 bp
* 30998 31711: contig of 714 bp in length
* 31712 31811: gap of 100 bp
* 31812 32528: contig of 717 bp in length
* 32529 32628: gap of 100 bp
* 32629 33340: contig of 712 bp in length
```

```
* 33341 33440: gap of 100 bp
* 33441 34135: contig of 695 bp in length
* 34136 34235: gap of 100 bp
* 34236 34937: contig of 702 bp in length
* 34938 35037: gap of 100 bp
* 35038 35758: contig of 721 bp in length
* 35759 35858: gap of 100 bp
* 35859 36547: contig of 689 bp in length
* 36548 36647: gap of 100 bp
* 36648 37345: contig of 698 bp in length
* 37346 37445: gap of 100 bp
* 37446 38186: contig of 741 bp in length
* 38187 38286: gap of 100 bp
* 38287 38996: contig of 710 bp in length
* 38997 39096: gap of 100 bp
* 39097 39787: contig of 691 bp in length
* 39788 39887: gap of 100 bp
* 39888 40608: contig of 721 bp in length
* 40609 40708: gap of 100 bp
* 40709 41434: contig of 726 bp in length
* 41435 41534: gap of 100 bp
* 41535 42253: contig of 719 bp in length
* 42254 42353: gap of 100 bp
* 42354 43056: contig of 703 bp in length
* 43057 43156: gap of 100 bp
* 43157 43866: contig of 710 bp in length
* 43867 43966: gap of 100 bp
* 43967 44675: contig of 709 bp in length
* 44676 44775: gap of 100 bp
* 44776 45483: contig of 708 bp in length
* 45484 45583: gap of 100 bp
* 45584 46311: contig of 728 bp in length
* 46312 46411: gap of 100 bp
* 46412 47129: contig of 718 bp in length
* 47130 47229: gap of 100 bp
* 47230 47941: contig of 712 bp in length
* 47942 48041: gap of 100 bp
* 48042 48766: contig of 725 bp in length
* 48767 48866: gap of 100 bp
* 48867 49592: contig of 726 bp in length
* 49593 49692: gap of 100 bp
* 49693 50404: contig of 712 bp in length
* 50405 50504: gap of 100 bp
* 50505 51202: contig of 698 bp in length
* 51203 51302: gap of 100 bp
* 51303 52008: contig of 706 bp in length
* 52009 52108: gap of 100 bp
* 52109 52810: contig of 702 bp in length
* 52811 52910: gap of 100 bp
* 52911 53642: contig of 732 bp in length
* 53643 53742: gap of 100 bp
* 53743 54459: contig of 717 bp in length
* 54460 54559: gap of 100 bp
* 54560 55267: contig of 708 bp in length
* 55268 55367: gap of 100 bp
* 55368 56077: contig of 710 bp in length
* 56078 56177: gap of 100 bp
```

Query Match 74.4%; Score 18.6; DB 78; Length 64231;  
Best Local Similarity 84.0%; Pred. No. 1.4e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 attgtcgaadtcacagaattattt 25  
||||||| |  
Db 32873 ATTGCGCAATGACAGAAATTCCTT 32849

## RESULT 41

AF128457  
LOCUS AF128457 70311 bp DNA PLN 25-APR-1999  
DEFINITION Oryza sativa subsp. indica BAC clone 16F19 php20725 region,  
complete sequence.  
ACCESSION AF128457

```

VERSION AF128457.1 GI:4680335
KEYWORDS
SOURCE
ORGANISM
Oryza sativa subsp. indica.
Oryza sativa subsp. indica
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
Oryza.
REFERENCE
1 (bases 1 to 70311)
AUTHORS Lou,A., Young,S. and Messing,J.
TITLE Microcollinearity in cereal genomes
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 70311)
AUTHORS Lou,A., Young,S. and Messing,J.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-1999) Waksman Institute, Rutgers University, 190
Frelinghuysen Rd., Piscataway, NJ 08854-8020, USA
FEATURES
source
1..70311
    /organism="Oryza sativa subsp. indica"
    /cultivar="Reging"
    /sub_species="indica"
    /db_xref="taxon:39946"
    /chromosome="10"
    /map="25.4"
    /clone="16r19"
    /note="php20725 region
    putative map location based on mapping of Oryza sativa EST
    clone E10310_62 found in GenBank Accession Number C99155"
    join(162..367,452..>556)
    /note="predicted gene - truncated (3'); similar to Oryza
    sativa EST clone 97AS2194 found in GenBank Accession
    Number AA753140"
    join(<2846..3003,4043..4325,4560..5015,5301..5833,
    6229..6424,6527..6685,6834..>7670)
    /product="hypothetical protein"
    join(2846..3003,4043..4325,4560..5015,5301..5833,
    6229..6424,6527..6685,6834..7670)
    /note="similar to Oryza sativa EST clones SL328_1A found
    in GenBank Accession Number D39744 and 96BS0679 found in
    GenBank Accession Number AA753512"
    /codon_start=1
    /product="hypothetical protein"
    /protein_id="AAD27627.1"
    /db_xref="GI:4680336"
    /translation="MDGEQEARWLAQGVAGADLVAAALRLQLEFFIAAVDRRLWLVGEG
    PLLERAIHREVLKMKIFQVOYLKDKRLYGRIDNSVSSIRAESKHQSEKVAEQ
    YPKPEFLENTSSDNIYANAGAAEDISYDLAAVKRQSFYQVDTPTHMDQRFLE
    EALARYGFLYLITNOENKMKLFRVPTDYDVVMMHSHQHPATYCHDMKLIGRVLE
    HDDDDDRSEKGLDTGSGTQEQFNAGARYKAGAMVGNLPSPTSNQPMFSGE
    VNGEFSVKAESQITILETVIEFLQIVDIKPLPAIPKENVYINFTKNQPMFISD
    GRLDIJSTKTKSIGASIOCEPTGELILTVLVDRTSSSKPKKIGKVSLSQFTWSD
    SKLSFERFELKPHDGHASSTPVSVRAASTVPVRAQQLVMIRTEPFLSKLSILSPN
    SVKQKMSQWTFYDCTNTELRIQIRKAKGMVVAELVGTVSKSKPKFLAEFV
    DNKWSLSNLCITNDMKPSKDGSLKELCDNRMKILYQGRLEFORKKCKNHAEEEDA
    SAITAVFSAEYPAVYKVALLDYKSELIMYKEDWFLPWTIVLSFMQDINVKDGEKLI
    LIGGMAQKDAISEPDTAAMATSAETVAAPANGCTGCTAGSNMAGDKVVAACRKA
    RCKAAGAGGSGTESAGCGGCGGCGGVAKVWEATKAGGCGGCGGGMWESSKAGHV
    GSGCGGCGGCGGMVYESSKDDYHAKSAGCGGCGGCGGCGGCGGGMWESSKAGHV
    KSGCGGCGGCGGCGGGMVYESSKDDYHAKSAGCGGCGGCGGCGGCGGGMWESSKAGHV
    GSGCGGCGGCGGCGGGMVYESSKDDYHAKSAGCGGCGGCGGCGGCGGGMWESSKAGHV
    7923..8069
    /rpt_family="putative MITE A"
    /rpt_type=dispersed
    8140..8386
    /rpt_type=dispersed
    10318..10483
    /rpt_family="putative MITE A"
    /rpt_type=dispersed
    join(<10691..10915,11490..11836,13787..>14354)
    /product="hypothetical protein"
    join(10691..10915,11490..11836,13787..14354)
    /codon_start=1
    /product="hypothetical protein"
    /protein_id="AAD27637.1"
    /db_xref="GI:4680346"
    /translation="MQLVSIPEQREPIKVIKTASADPARLLQLARHPSPSTLHRTA
    PHRALATITARRADPAEGRKSRVAPRAPRAGVLPTGLFLAVGAWHVAARAAAD
    PGRFLRANPVDVGGGAPAMLPAHLEYIYAGGAFHMCVEVLYSTHLHIFADGGI
    NPAHLNDLHGGMMLFFLGLALLSQKRYLPLPEGALCLVASTAFMAELLFYFSTTHQLEGYYH
    STTHQLEGYYHYLLVVVVALCVATTVLGLALLPASFPVDIAGSAAIALQGLWFTYQTAFTLYGSLPAGCR
    TLGSPSPAGCRDADGHDCHTHAAGAEOLANFOLGLVFLVCAVALGCEFAVAAAHGHPDLATMHA
    RHGHPDLATMHAHVAAMBAQALAGACAGGDRFVGSALPLEDTAI"
    join(<11475..11836,13787..>14354)
    /product="hypothetical protein"
    join(11475..11836,13787..14354)
    /codon_start=1
    /product="hypothetical protein"
    /protein_id="AAD27628.1"
    /db_xref="GI:4680337"
    /translation="MGSPKGVHLPGLTFLAVGAWHVAARAAADPGRFLRANPVD
    DVGCGAPAWLPAHLEYIYAGGAFHMCVEVLYSTHLHIFADGGINPAHLNDLHG
    YLIMFFLGLALLSQKRYLPLPEGALCLVASTAFMAELLFYFSTTHQLEGYYH
    YLLVVVVALCVATTVLGLALLPASFPVDIAGSAAIALQGLWFTYQTAFTLYGSLPAGCR
    RDADGHDCHTHAAGAEOLANFOLGLVFLVCAVALGCEFAVAAAHGHPDLATMHA
    KHVAAMBAQALAGACAGGDRFVGSALPLEDTAI"
    join(<16072..16311,16316..>16834)
    /product="hypothetical protein"
    join(16072..16311,16316..16834)
    /codon_start=1
    /product="hypothetical protein"
    /protein_id="AAD27635.1"
    /db_xref="GI:4680344"
    /translation="MWASWSTDECAKRRTRRGQRGVRGGGGGGEAGKLLLSNPS
    PHLTPPTREEGGGGAGVGTGMEHHLACTAYRTPAMETMSVVAAATIIIGIMRHPSARL
    SRHIFPYSNMYTAGIVLHYHAPLRPNRYAVPTPLHYQLRSTTHVPTTPLYHAY
    LRSWMLTLRGTHVMDLPSLSPPLRLVAHRTTSLVPAVPPHSAWLVNVAQPDH
    RHRRGEGPSPSPRCSSTSPSPSSGILLTVEKR"
    19801..19945
    /rpt_family="putative MITE B"
    /rpt_type=dispersed
    20325..20653
    /rpt_type=dispersed
    complement(20546..20624)
    /rpt_family="putative MITE B"
    21168..21288
    /rpt_family="putative MITE H"
    /rpt_type=dispersed
    24167..24475
    /rpt_type=dispersed
    24333..24438
    /rpt_family="putative MITE B"
    /rpt_type=dispersed
    complement(join(<25162..25820,26472..>27849))
    /product="putative receptor kinase"
    complement(join(25162..25820,26472..27849))
    /note="similar to Oryza sativa EST clone E2886_2A found in
    GenBank Accession Number C73087"
    /codon_start=1
    /product="putative receptor kinase"
    /protein_id="AAD27636.1"
    /db_xref="GI:4680345"
    /translation="MVTLLAFRLSTLLLLLAAGATAVDPDGAAAPDPTAAAPDTEA
    AALLRLKSLIDPTNALEAWSPSPSPPCDEAHRWRVQCYNGVILGLRLARLNSGD
    FDFALSRPLGLHSILIRNNPSPPLPASLAARSLRALYLSRNAFGSPVPGDFAAM
    SWLKLYLDNRNFGSELPAALAGAPRLOELHLDNRIEGRVPSKLPALPLFNVSHN
    RLTVLPFAVAAARFNESAFNPGCLGAPGAGCAAAAACGPAHSGHNPMSAADIYA
    VOETSFVFMGIMLVLLVAVAGVLMRLRDEGTSTASSGHEHPAIGAPSGNLSVPH
    AAGAAASQLVMTMEGGGSGGSGGVMGKRVQAEFVLSNAAGELRPLMLKASAEVL
    NGTLGSAYKAAMRNGTVAVVRMGRNRVGRFEEHIRMELGRLRTPSTSPSATIT
    ARKSSSPSDQSDRVLVDWPMARIVAVVGRVLSYHKEGLIPAMRLVMTGADFD
    APPPPHNLKSGNILLDALEPRTIVDYGFFLVNTSOAPHAMFRSPSAASAAGA
    GAGAAOAAALSARSDDYICLVILLELVTKFEPQYLLTARGGTDVQVWAAASVAGGT
    EQEVDPVVAAGAPAAVRLLRVGRCTIPEPESRNPMSADVARMVQVAGGAS"
    35340..35604
    /stem_loop
    repeat_region
  
```





```

repeat_region 11062..11292
/note="MIR repeat: matches 20. .262 of consensus"
repeat_region 11626..11639
/note="MIR repeat: matches 249. .262 of consensus"
repeat_region 11640..11672
/note="L2 repeat: matches 2660. .2692 of consensus"
repeat_region 11673..11689
/note="MIR repeat: matches 199. .250 of consensus"
repeat_region 11925..12015
/note="MIR repeat: matches 63. .183 of consensus"
repeat_region 12196..12339
/note="MIR repeat: matches 73. .212 of consensus"
repeat_region 12804..13058
/note="MIR repeat: matches 5. .262 of consensus"
repeat_region 14554..14790
/note="MIR repeat: matches 4. .251 of consensus"
repeat_region 15333..15631
/note="AluY repeat: matches 1. .299 of consensus"
repeat_region 15991..16210
/note="Tigger3(Golem) repeat: matches 2805. .3023 of consensus"
repeat_region 16273..16541
/note="L1MA6 repeat: matches 5660. .5939 of consensus"
repeat_region 16542..16705
/note="Tigger3(Golem) repeat: matches 2195. .2362 of consensus"
repeat_region 16706..17031
/note="L1MA6 repeat: matches 5951. .6294 of consensus"
repeat_region 17152..17242
/note="L1ME3A repeat: matches 6034. .6121 of consensus"
repeat_region 19105..19171
/note="MIR repeat: matches 156. .222 of consensus"
repeat_region 19489..19617
/note="MIR repeat: matches 58. .210 of consensus"
repeat_region 20380..20419
/note="20 copies 2 mer aa 80 conserved"
repeat_region 20593..20710
/note="L2 repeat: matches 2592. .2710 of consensus"
repeat_region 21342..21654
/note="AluY repeat: matches 3. .310 of consensus"
repeat_region 22842..22981
/note="L2 repeat: matches 2033. .2168 of consensus"
repeat_region 23099..23546
/note="L1A1 repeat: matches 1. .501 of consensus"
repeat_region 23652..23748
/note="L2 repeat: matches 1674. .1761 of consensus"
repeat_region 23787..26374
/note="L1MA1 repeat: matches 2962. .5632 of consensus"
repeat_region 26375..26678
/note="AluSg repeat: matches 1. .299 of consensus"
repeat_region 26679..27359
/note="L1MA1 repeat: matches 5632. .6300 of consensus"
repeat_region 27865..28056
/note="L2 repeat: matches 1332. .1536 of consensus"
repeat_region 28064..28229
/note="L1MA repeat: matches 2839. .3022 of consensus"
repeat_region 28253..28383
/note="AluJo repeat: matches 3. .134 of consensus"
repeat_region 28392..28673
/note="AluJo repeat: matches 2. .301 of consensus"
repeat_region 28758..28847
/note="45 copies 2 mer ag 62 conserved"
repeat_region 28851..28930
/note="40 copies 2 mer ag 63 conserved"
repeat_region 29669..29965
/note="AluJo repeat: matches 1. .303 of consensus"
gene complement(30093..30600)
/gene="CB42E1.1"
CDS complement(<30093..>30600)
/gene="CB42E1.1"
/note="supported by FGENES and GENSCAN; match: cDNAs;
Em:U54999; match: proteins: Tr:P81274 Tr:O58663 Tr:O55489
Tr:O28309 Tr:O42393 Sw:Q13702 Sw:P12672 Sw:P09108

```

```

Tr:O26287 Wp:CE04472"
/codon_start=1
/evidence=not_experimental
/product="dA45p1.1 (PUTATIVE partial novel protein with
remote similarity to RAPSIN and LGN)"
/protein_id="CAB62975.1"
/db_xref="GI:6572245"
/translation="AASSALSSLGHHVYTAIGDYPNALASHKOCVLLAKOSKDELSEAR
ELNGMGAVIAMGDFENAVOCHEQHLKIAKDLGNKEPEARAYSNLGSAYHYRNPDKA
MSYHNYVLELAQELMEKALEMAYAGLGHAAQMDLERAKOYHEQUGIAEDLKDRA
AGRASSNL"
repeat_region 30685..306823
/note="L1PA2 repeat: matches 12. .6144 of consensus"
repeat_region 37207..37517
/note="AluY repeat: matches 1. .309 of consensus"
misc_feature complement(38024..38251)
/note="match: GSS: Em:B14059"
repeat_region 39179..39267
/note="MIR repeat: matches 34. .128 of consensus"
repeat_region 39844..39979
/note="MIR repeat: matches 77. .212 of consensus"
repeat_region 41186..41330
/note="MIR repeat: matches 105. .260 of consensus"
repeat_region 41579..41739
/note="AluJo/FRAM repeat: matches 148. .306 of consensus"
repeat_region 41946..42450
/note="L1MA5A repeat: matches 4820. .5325 of consensus"
repeat_region 42451..42751
/note="AluSg repeat: matches 1. .298 of consensus"
repeat_region 42752..43719
/note="L1MA5A repeat: matches 5325. .6295 of consensus"
repeat_region 43745..45171
/note="L1MC2 repeat: matches 4830. .6324 of consensus"
repeat_region 45174..45466
/note="AluSg repeat: matches 1. .281 of consensus"
repeat_region 45529..45556
/note="L1MC4 repeat: matches 6741. .6768 of consensus"
repeat_region 45716..45918
/note="MIR repeat: matches 20. .212 of consensus"
repeat_region 45970..46013
/note="L2 repeat: matches 2157. .2203 of consensus"
repeat_region 46148..46460
/note="AluSc repeat: matches 1. .309 of consensus"

Query Match 74.4%; Score 18.6; DB 92; Length 77601;
Best Local Similarity 84.0%; Pred. No. 1.4e-02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25
||||| ||||| ||||| ||||| |||||
Db 16455 ATGTCTTAAGTCACAGAAATTTT 16431

RESULT 43
ATM4E13 80346 bp DNA PLN 07-MAY-1999
LOCUS Arabidopsis thaliana DNA chromosome 4, BAC clone M4E13 (ESSAI
DEFINITION project).
ACCESSION AL022023
VERSION AL022023.1 GI:2924505
KEYWORDS thale cress.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 80346)
AUTHORS Bevan,M., Purnelle,B., Boutry,M., Goffeau,A., Hoheisel,J.,
Mewes,H.W., Mayer,K.F.X. and Schueller,C.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 80346)
AUTHORS EU Arabidopsis sequencing project.
TITLE Direct Submission

```





\* 16308 17086: contig of 779 bp in length  
 \* 17087 17186: gap of 100 bp  
 \* 17187 18005: contig of 819 bp in length  
 \* 18006 18105: gap of 100 bp  
 \* 18106 18928: contig of 823 bp in length  
 \* 18929 19028: gap of 100 bp  
 \* 19029 19847: contig of 819 bp in length  
 \* 19848 19947: gap of 100 bp  
 \* 19948 20753: contig of 806 bp in length  
 \* 20754 20853: gap of 100 bp  
 \* 20854 21657: contig of 804 bp in length  
 \* 21658 21757: gap of 100 bp  
 \* 21758 22551: contig of 794 bp in length  
 \* 22552 22651: gap of 100 bp  
 \* 22652 23474: contig of 823 bp in length  
 \* 23475 23574: gap of 100 bp  
 \* 23575 24389: contig of 815 bp in length  
 \* 24390 24489: gap of 100 bp  
 \* 24490 25281: contig of 792 bp in length  
 \* 25282 25381: gap of 100 bp  
 \* 25382 26168: contig of 787 bp in length  
 \* 26169 26268: gap of 100 bp  
 \* 26269 26927: contig of 659 bp in length  
 \* 26928 27027: gap of 100 bp  
 \* 27028 27845: contig of 818 bp in length  
 \* 27846 27945: gap of 100 bp  
 \* 27946 28772: contig of 827 bp in length  
 \* 28773 28872: gap of 100 bp  
 \* 28873 29688: contig of 816 bp in length  
 \* 29689 29788: gap of 100 bp  
 \* 29789 30607: contig of 819 bp in length  
 \* 30608 30707: gap of 100 bp  
 \* 30708 31491: contig of 784 bp in length  
 \* 31492 31591: gap of 100 bp  
 \* 31592 32396: contig of 805 bp in length  
 \* 32397 32496: gap of 100 bp  
 \* 32497 33284: contig of 788 bp in length  
 \* 33285 33384: gap of 100 bp  
 \* 33385 34191: contig of 807 bp in length  
 \* 34192 34291: gap of 100 bp  
 \* 34292 35087: contig of 796 bp in length  
 \* 35088 35187: gap of 100 bp  
 \* 35188 35998: contig of 812 bp in length  
 \* 36000 36098: gap of 100 bp  
 \* 36100 36944: contig of 845 bp in length  
 \* 36945 37044: gap of 100 bp  
 \* 37045 37862: contig of 818 bp in length  
 \* 37863 37962: gap of 100 bp  
 \* 37963 38791: contig of 829 bp in length  
 \* 38792 38891: gap of 100 bp  
 \* 38892 39719: contig of 828 bp in length  
 \* 39720 39819: gap of 100 bp  
 \* 39820 40615: contig of 796 bp in length  
 \* 40616 40715: gap of 100 bp  
 \* 40716 41525: contig of 810 bp in length  
 \* 41526 41625: gap of 100 bp  
 \* 41626 42421: contig of 796 bp in length  
 \* 42422 42521: gap of 100 bp  
 \* 42522 43314: contig of 793 bp in length  
 \* 43315 43414: gap of 100 bp  
 \* 43415 44203: contig of 789 bp in length  
 \* 44204 44303: gap of 100 bp  
 \* 44304 45119: contig of 816 bp in length  
 \* 45120 45219: gap of 100 bp  
 \* 45220 46020: contig of 801 bp in length  
 \* 46021 46120: gap of 100 bp  
 \* 46121 46891: contig of 771 bp in length  
 \* 46892 46991: gap of 100 bp  
 \* 46992 47789: contig of 798 bp in length  
 \* 47790 47889: gap of 100 bp  
 \* 47890 48705: contig of 816 bp in length  
 \* 48706 48805: gap of 100 bp  
 \* 48806 49633: contig of 828 bp in length

\* 49634 49733: gap of 100 bp  
 \* 49734 50556: contig of 823 bp in length  
 \* 50557 50656: gap of 100 bp  
 \* 50657 51474: contig of 818 bp in length  
 \* 51475 51574: gap of 100 bp  
 \* 51575 52363: contig of 789 bp in length  
 \* 52364 52463: gap of 100 bp  
 \* 52464 53248: contig of 785 bp in length  
 \* 53249 53348: gap of 100 bp  
 \* 53349 54133: contig of 785 bp in length  
 \* 54134 54233: gap of 100 bp  
 \* 54234 55027: contig of 794 bp in length  
 \* 55028 55127: gap of 100 bp  
 \* 55128 55928: contig of 801 bp in length  
 \* 55929 56028: gap of 100 bp  
 \* 56029 56820: contig of 792 bp in length  
 \* 56821 56920: gap of 100 bp  
 \* 56921 57725: contig of 805 bp in length  
 \* 57726 57825: gap of 100 bp  
 \* 57826 58638: contig of 813 bp in length  
 \* 58639 58738: gap of 100 bp  
 \* 58739 59520: contig of 782 bp in length  
 \* 59521 59620: gap of 100 bp  
 \* 59621 60439: contig of 819 bp in length  
 \* 60440 60539: gap of 100 bp  
 \* 60540 61374: contig of 835 bp in length  
 \* 61375 61474: gap of 100 bp  
 \* 61475 62272: contig of 798 bp in length  
 \* 62273 62372: gap of 100 bp

Query Match 74.4%; Score 18.6; DB 68; Length 82189;

Best Local Similarity 84.0%; Pred. No. 1.4e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 atttgcgaagtcacagaattattt 25

||||| |||||||||||||||

Db 48302 ATTCTTTCATGTCACAGAAATTATT 48326

# RESULT 45

ATT12J5

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

24-FEB-1999

T12J5 (ESSA11

project).

AL035522

AL035522.1

GI:4455339

thale cress.

Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 84499)

Bevan, M., Murphy, G., Ridley, P., Hudson, S., Hoheisel, J., Mewes, H.W.,

Mayer, K.F.X. and Schueller, C.

Unpublished

2 (bases 1 to 84499)

EU Arabidopsis sequencing project.

Direct Submission

Submitted (22-FEB-1999) MIPS, at the Max-Planck-Institut fuer

Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:

schueller@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project

Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge

Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,

E-mail: michael.bevan@bbsrc.ac.uk

Information on performance of analysis and a more detailed

annotation of this entry and other sequences of chromosome 4 can be

viewed at: <http://websvr.mips.biochem.mpg.de/proj/thal/>.

Location/Qualifiers

1. .84499

/organism="Arabidopsis thaliana"

/variety="Columbia"

/db\_xref="taxon:3702"

FEATURES

SOURCE

```

/misc_feature
1..49380
/notes="position 1-49380 overlaps to BAC clone M4E13, EMBL
acc:AL022023; for sequence analysis and annotation please
refer to this entry"
48992..51335
/gene="T12J5.10"
/gene="T12J5.10"
complement(join(48992..49360,49762..50493,50815..51066,
51207..51335))
/gene="T12J5.10"
complement(join(48992..49360,49762..50493,50815..51066,
51207..51335))
/gene="T12J5.10"
/notes="similarity to Human (H326) mRNA, Homo sapiens,
gb:U06631"
/codon_start=1
/product="putative protein"
/protein_id="CAB36721.1"
/db_xref="GI:4455340"
/translation="MSEKRRGRSNGFLNHPVVDVFWREVGIGSSRNFSNRFSASEN
LVLRLEIKLEKHKGCYNTVSFNAEGDVLISGSDRRVVLMDQLGNVKLSFHSCHA
NNVFAKMPFSDRTIVTCAADGFDLRTAEPTLFTCRSDVPRRRNMDAIOINATA
IDPRSNLFAVGMEYEARLYDIRRFOGELNGPTRAADHFCPPHLIGNEVDGITGLA
FQEQLLVSYNDEPIVLTBQMGILGSPIPSSIPKSPVSESSSPKDNENHSVC
LVYKHKCEVKGVPFGRSEYVVSQDGRIFIMRKGGLIRIWEADRHVVNCI
EPHPIPLVASSGESIDIKVWTSKAAERATIPENIELLPSPRIPWLSFSFHDYDDE
LFGNMDIGIDGNEGEDSIDDAEDDDDDSSGVLDLDDNDSDDDDDSDDDDDDD
DSDDFCFEIDNNNDNNGDGGSETNVDSISGSHQDDDDDD"
complement(48992..49360)
/gene="T12J5.10"
/number=1
complement(49361..49761)
/number=1
complement(49762..50493)
/gene="T12J5.10"
/number=2
complement(50494..50814)
/number=2
complement(50815..51066)
/gene="T12J5.10"
/number=3
complement(51067..51206)
/number=3
complement(51207..51335)
/gene="T12J5.10"
/number=4
complement(join(51806..52129,52611..52939,53065..53389))
/gene="T12J5.20"
complement(51806..52129)
/gene="T12J5.20"
/number=1
51806..53389
/gene="T12J5.20"
complement(join(51806..52129,52611..52939,53065..53389))
/gene="T12J5.20"
/notes="strong similarity to caffeic acid
O-methyltransferase, Pinus taeda, gb:U39301"
/codon_start=1
/product="O-methyltransferase-like protein"
/protein_id="CAB36722.1"
/db_xref="GI:4455341"
/translation="MEESKRNLLDEEAKASLDIWRVYVGFADIAAKCATDLKIPAI
ENHPSSQPVTLSELSSAVSAPSHLRIRMEFLVHQLKEVPTKDLGATGTYNTPLSR
RMTITKLHGDLLAFQADNLCHSQLNEAMACDARRVPRVAGACGLGFDVATVVDV
GGGTGTGILVKEFPWKGFNEDLPHVIEVAQVLVDGEMDFDIPASDAVLIK
WLLHWDGDKCIIKLKNCRAVPNIGKVLIVCEVIGERKNTMIAERDDKLEHVRLO
LDMMVMVSTGKERTLKWDVFLTEAGFARYEVRDIDDVQSLIIAYS"
complement(52130..52610)
/number=1
complement(52611..52939)
/gene="T12J5.20"
/number=2
complement(52940..53064)
/number=2
complement(53065..53389)
/gene="T12J5.20"
/number=3
55847..57666
/gene="T12J5.30"
complement(join(55847..56170,56842..57666))
/gene="T12J5.30"
complement(join(55847..56170,56842..57666))
/notes="strong similarity to caffeic acid
O-methyltransferase, Pinus taeda, gb:U39301
contains EST gb:N96576, W43582"
/codon_start=1
/product="O-methyltransferase-like protein"
/protein_id="CAB36723.1"
/db_xref="GI:4455342"
/translation="MSSDQISKFLDRNMEDNKRKVLDEEAKASLDIWRVYVGFADIA
AAKCATDLKIPAIENHPSSQPVTLAEELSSAVSAPSHLRIRMEFLVHQLKEIPTK
DGLATGVNTPLSRRLMITRRDCKSLAPFLVETPEMLAPMLRSSVSVSSVNGSTP
PPDVAHGVKDVNSFAODNBPFLSDMINEMACDARRVPRVACAGHGLFGCVTTMDVG
GGTGTGMLVREFPWIKEFNPDLPVIEVAELDGVENVEGDMFDSIPACDAIFIKW
VLHWDGDKCIIKLKNCRAVPNIGKVLIVCEVIGERKNTMIVDERDDKLEHVRML
DVMYMAHSTGKERTLKWDVFLTEAGFARYEVRDIDDVQSLIIAYS"
complement(55847..56170)
/gene="T12J5.30"
/number=1
complement(56171..56841)
/number=1
complement(56842..57666)
/gene="T12J5.30"
/number=2
61697..62188
/gene="T12J5.40"
61697..62188
/gene="T12J5.40"
/number=1
61697..62188
/gene="T12J5.40"
/notes="similarity to predicted protein, Arabidopsis
thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAB36724.1"
/db_xref="GI:4455343"
/translation="MMTWPARSSPONTKRPVYVHVPPNVDVKITSGSGSPGSPGL
NDQGOVSNPQHHSVAESSYPRSSGPLRNEYSSVOVHDLDRTHEDEYDEMDDGPEK
RRIITRFVSCLLFTVLVLAFTLFCLLILWGSVSFAPIATLKVIDREDRLRENFTFISV
FKI"
complement(join(63375..63860,63950..64081,64151..64551,
64736..64869,64975..65243))
/gene="T12J5.50"
complement(63375..63860)
/gene="T12J5.50"
/number=1
63375..65243
/gene="T12J5.50"
complement(join(63375..63860,63950..64081,64151..64551,
64736..64869,64975..65243))
/gene="T12J5.50"
/notes="strong similarity to Lily mRNA, Lillium longiflorum,
gb:D21814
Contains prs HPR component phosphorylation sites
signatures, Pts_Hpr_Ser [SQFPLNLSLFGVSLIG]"
/codon_start=1
/product="amino acid permease-like protein"
/protein_id="CAB36725.1"
/db_xref="GI:4455344"
/translation="MSIALGNLFDEESGSGSPLFMSAPSTDQPISEKNGDGG
RIPVEEWLPITESRKNAGNYTATFHLCSGIGLQVILLPAFPAALGWMTGTLITVGFV
WKLYTLLVQLHEAVFGIRISRYVRLAIFSGGACTILVITGKSIQQLIQIMSDDN
TAPLTSVQCFLVSCIAMISQFPLNLSLFGVSLIGAFMGITAYCTVITWILPVASDSOR
TQVSYSVATMDKSEVHFNAIGLIALYVRGNLVLEIQVLNLSQLKHLDDGLTSPDSK

```

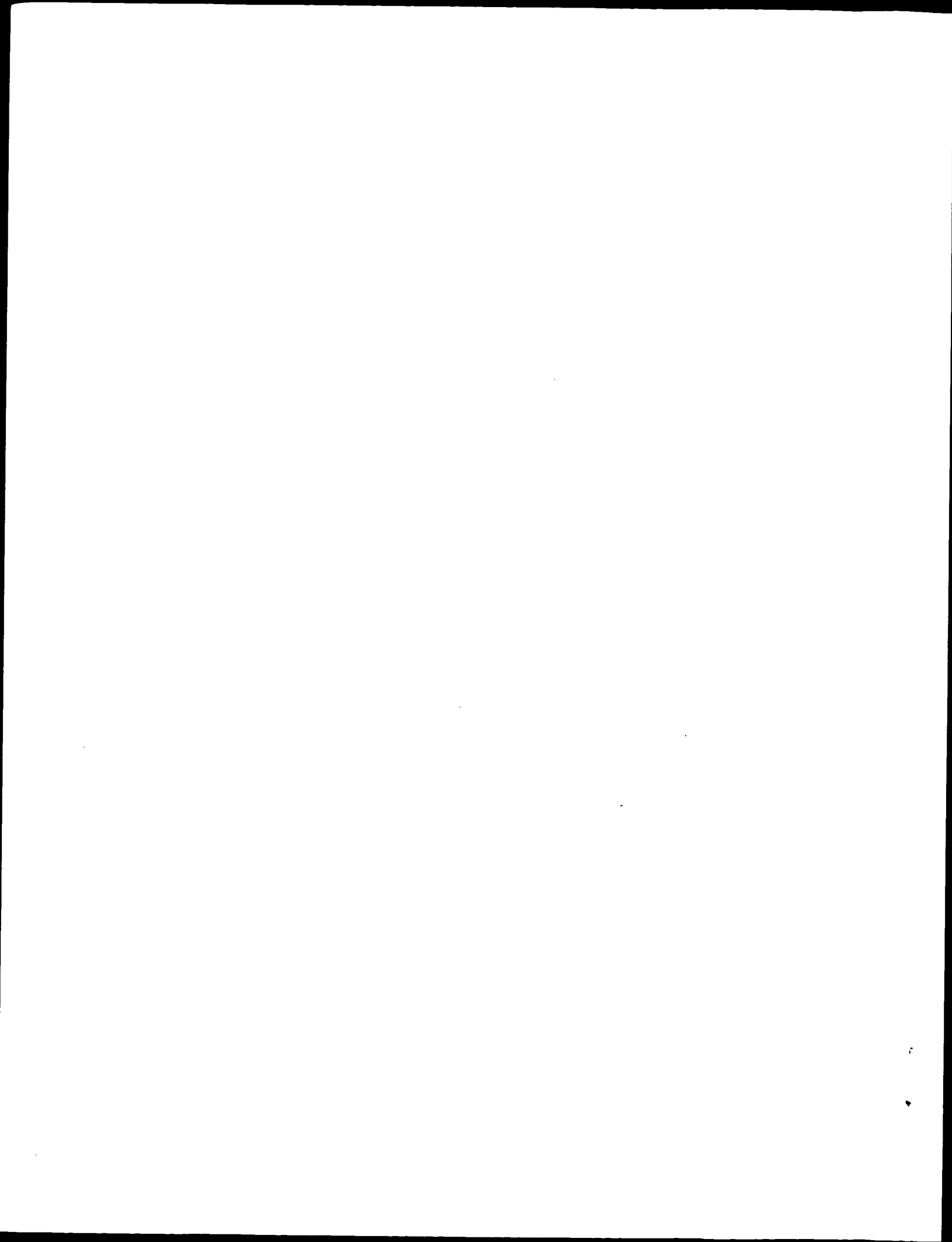
NPSCKTMRAVNHSHALVAICMFPLTFAYWAYGDKIPATGSPVGNLYLKLYTOEHSKR  
AACFTHLTFIFSCLSYPINLMPACDNIEMVYITKKKKPASIIVRMMLRVFLSLVCFT  
IAGPPFLPYLAVLIGATLALVTITYPCFEMWISIKKPKQKSPMWLFNVLVGLGASLS  
VLLLVASAMRLAQKGLHANFRP"

intron complement(63861. .63949)  
exon /number=1  
complement(63950. .64081)  
/gene="T12J5.50"  
intron /number=2  
complement(64082. .64150)  
/number=2  
complement(64151. .64551)  
exon /gene="T12J5.50"  
/number=3  
intron complement(64552. .64735)  
/number=3  
complement(64736. .64869)  
exon /gene="T12J5.50"  
/number=4  
intron complement(64870. .64974)  
/number=4  
exon complement(64975. .65243)

Query Match 74.4%; Score 18.6; DB 13; Length 84499;  
Best Local Similarity 84.0%; Pred. No. 1.4e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 attgtcgaagtcacagaattattt 25  
|||||  
Db 36816 ATTGCGCAATAACAGATTGATTT 36840

Search completed: October 9, 2001, 12:09:34  
Job time: 3689 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:35:54 ; Search time 470.56 Seconds  
(without alignments)  
33.359 Million cell updates/sec

```

Title: US-09-396-196F-2
Perfect score: 25
Sequence: 1 attgtcgaagtccacagaattattt 25

```

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues  
Total number of hits satisfying chosen parameters: 1460202

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database : N_Geneseq_0601.*
1: /SIDSL/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqn/NA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query %		DB	ID	Description
		Match	Length			
1	25	100.0	1041	20	AA01303	E. coli biotin syn
2	25	100.0	1084	10	AA01329	E.coli Bio B gene.
3	25	100.0	1121	7	AA060496	Sequence encoding
4	25	100.0	5872	15	AA062386	Biotin-biosynthesi
c 5	18.6	74.4	902	21	AAZ56756	Human transmembran
6	18.6	74.4	2114	14	AAQ51236	Plant NADPH cytoch
c 7	17.6	71.2	1038602	20	AAZ01425	Complete genome se
8	17.6	70.4	564	21	AAC44922	Arabidopsis thalia
9	17.6	70.4	3881	21	AA01269	Human secreted pro
c 10	17.2	68.8	386	21	AAC24233	Human secreted pro
11	17.2	68.8	13830	22	AAD02659	Tomato chromosome

c	12	17	68.0	17	14	AAQ56307
c	13	17	68.0	512	20	AAV88744
c	14	17	68.0	4597	19	AAV52308
c	15	17	68.0	10095	14	AAQ11762
c	16	16.8	67.2	3083	12	AAQ11851
c	17	16.6	66.4	274	16	AAQ121899
c	18	16.6	66.4	322	16	AAQ191339
c	19	16.6	66.4	1041	10	AAQ191327
c	20	16.6	66.4	2730	21	AAQ760297
c	21	16.6	66.4	32351	21	AAQ213029
c	22	16.6	66.4	32351	21	AAQ335185
c	23	16.6	66.4	40298	21	AAQ213311
c	24	16.6	66.4	40298	21	AAQ335189
c	25	16.2	64.8	500	18	AAQ343451
c	26	16.2	64.8	1001	21	AAQ57499
c	27	16.2	64.8	1001	21	AAQ575400
c	28	16.2	64.8	1002	21	AAQ575945
c	29	16.2	64.8	1374	15	AAQ68434
c	30	16.2	64.8	1374	16	AAQ69463
c	31	16.2	64.8	1771	20	AAQ30602
c	32	16.2	64.8	5198	15	AAQ68433
c	33	16.2	64.8	5198	16	AAQ93462
c	34	16.2	64.8	8341	18	AAQ68772
c	35	16.2	64.8	8767	18	AAQ68735
c	36	16.2	64.8	8768	18	AAQ73812
c	37	16.2	64.8	8799	18	AAQ68758
c	38	16.2	64.8	8816	18	AAQ68750
c	39	16.2	64.8	8873	18	AAQ68740
c	40	16.2	64.8	8970	18	AAQ68741
c	41	16.2	64.8	8996	18	AAQ68771
c	42	16.2	64.8	8996	18	AAQ68783
c	43	16.2	64.8	8997	18	AAQ68754
c	44	16.2	64.8	8997	18	AAQ68774
c	45	16.2	64.8	9006	18	AAQ68756
c	46	16.2	64.8	9006	18	AAQ68775

## ALIGNMENTS

## RESULT

RESOLUT L  
AAX01303  
ID AAX01303 standard; DNA; 1041 BP.

AA  
AC AAX01303;

12-APR-1999 (first entry)

E. coli biotin synthetase (BioB) coding sequence.

DAP aminotransferase; diaminopelargonic acid; transgenic plant;  
KW biotin synthase; biotin production; vitamin H; BtOS; ss.

OS Escherichia coli.

XX  
PN  
US5869719-A

XX  
PD 09-FEB-1999

30-APR-1997; 97US-0846338.

30-APR-1997: 97US-0846338

08-MAR-1995; 95US-0401068.  
XX

PA (NOVS ) NOVARTIS FINANCE CORP.  
XX

Patton DA;

WPI; 1999-152902/13.  
P-PSDB: AAW73906

Transgenic plants

encoding di:amino-pelargonic acid amino-transferase or biotin synthase

XX PS Example 2; Column 37-40; 34pp; English.

XX CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can

XX CC be used in the transgenic plant of the invention. The transgenic plant,

XX CC plant cell or plant tissue is transformed with a chimeric gene encoding

XX CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and

XX CC produces more biotin than a non-transgenic plant, cell or tissue. The

XX CC plant is used as an improved dietary source of biotin (vitamin H) for

XX CC humans or animals.

XX SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;

Best Local Similarity 100.0%; Pred. No. 0.011;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25

Db 24 attgtcgcaagtcacagaattattt 48

RESULT 2

AAAN91329

ID AAAN91329 standard; DNA; 1084 BP.

XX AC AAAN91329;

XX DT 15-FEB-1990 (first entry)

XX DE E.coli Bio B gene.

XX KW E.coli; Bio B gene; biotin.

XX OS Escherichia coli.

XX FH Key Location/Qualifiers

XX FT CDS 24..1064

XX FT /\*tag=a

XX GN GB2216530-A.

XX PD 11-OCT-1989.

XX PF 17-MAR-1989; 89GB-0006210.

XX PR 22-MAR-1988; 88GB-0006804.

XX PR 17-MAR-1989; 89GB-0006210.

XX PA (UKAG-) UK MIN. AGRIC. FISH.

XX PI Pearson BM, McKee RA;

XX DR WPI; 1989-295085/41. P-PSDB P91392

XX PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes

XX PT - derived from E.coli and capable of replication and expression in other

XX PT microorganisms, esp. yeast.

XX PS Table 3; page 33-4; 52pp; English.

XX CC The gene can be used in a plasmid for expression of enzymes of the biotin

XX CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae

XX CC are plasmids pMA91, pKVA36c, pKVA49 and pCK495, and plasmid pCK965 for

XX CC Lactobacillus. Insertion of bio B improves biotin yields in

XX CC microorganisms which export biotin, or enables growth in media contg.

XX CC little or no biotin of organisms unable to synthesise biotin for their

XX CC own use.

XX SQ Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;

Best Local Similarity 100.0%; Pred. No. 0.011;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25

Db 47 attgtcgcaagtcacagaattattt 71

RESULT 3

AAAN60496

ID AAAN60496 standard; DNA; 1121 BP.

XX AC AAAN60496;

XX DT 17-OCT-1991 (first entry)

XX DE Sequence encoding biotin synthesising enzyme.

XX KW Biotin synthetic enzyme; E.coli; desthiobiotin; ds.

XX FH Key Location/Qualifiers

XX FT CDS 42..1082

XX FT /\*tag= a

XX GN JP61149091-A.

XX PD 07-JUL-1986.

XX PF 24-DEC-1984; 84JP-0272605.

XX PR 24-DEC-1984; 84JP-0272605.

XX PA (NIPS ) NIPPON SODA KK.

XX WPI; 1986-216622/33.

XX P-PSDB; AAP60536.

XX Double stranded DNA encoding biotin synthesising enzyme -

XX PT comprises transformed mutant E.coli strain contg. cyclic doubled

XX PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.

XX PS Disclosure; Page 534; 23pp; Japanese.

XX CC The sequence may be expressed by a transformed E.coli host, cultured

XX CC in a medium containing desthiobiotin.

XX SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;

Best Local Similarity 100.0%; Pred. No. 0.012;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25

Db 65 attgtcgcaagtcacagaattattt 89

RESULT 4

AAQ62386

ID AAQ62386 standard; DNA; 5872 BP.

XX AC AAQ62386;

XX DT 16-NOV-1994 (first entry)

XX DE Biotin-biosynthesis genes contg. plasmid pB030A-15/9.

XX KW Biotin; expression; enterobacteria; vitamin H; synthesis;

XX KW plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;

XX KW promoter ptac; biotin synthase; KAPA synthase;

XX KW 8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTP synthase;

KW dethiobiotin synthase; DAPA synthase;  
 KW S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;  
 KW seborrhea; dermatitis; ds.  
 XX  
 OS Escherichia coli DSM498.

```

Key      Location/Qualifiers
promoter 1..96
          /tag= a
          /function= "promoter ptac"
          /evidence= EXPERIMENTAL
-35_signal 23..28
          /tag= b
          /standard_name= "promoter ptac"
-10_signal 45..50
          /tag= c
          /evidence= EXPERIMENTAL
          /standard_name= "promoter ptac"
RBS      105..109
          /tag= d
          /evidence= EXPERIMENTAL
          /standard_name= "bioB RBS no. 9"
CDS      117..1157
          /tag= e
          /product= "biotin synthase"
          /evidence= EXPERIMENTAL
          /gene= "bioB"
          /number= 1
          /tag= f
          /standard_name= "bioF RBS"
          /tag= g
          /EC_number= 2.3.1.47
          /product= "RAPA synthase"
          /evidence= EXPERIMENTAL
          /gene= "bioF"
          /number= 2
          /standard_name= "8-amino-7-oxononanoate synthase"
          /tag= h
          /standard_name= "bioC RBS"
          /tag= i
          /number= 3
          /standard_name= "bioD RBS"
          /tag= j
          /number= 4
          /EC_number= 6.3.3.3
          /product= "DTB synthase"
          /evidence= EXPERIMENTAL
          /gene= "bioD15"
          /number= 4
          /standard_name= "dethiobiotin synthase"
          /tag= l
          /note= "bioD15 substitution"
          /tag= m
          /standard_name= "bioA RBS"
          /tag= n
          /EC_number= 2.6.1.62
          /product= "DAPA synthase"
          /evidence= EXPERIMENTAL
          /gene= "bioA"
          /number= 5
          /standard_name= "S-adenosyl-L-methionine: 8-amino-
```

```

FT RBS      5088..5093
FT      /tag= o
FT      /standard_name= "ORFI RBS"
FT CDS      5098..5574
FT      /tag= p
FT      /function= "unknown, involved in biotin synthesis"
FT      /product= "protein"
FT      /evidence= EXPERIMENTAL
FT      /gene= "ORFI"
FT      /number= 6
FT      /tag= q
FT      /standard_name= "rho-independent transcriptional
FT      terminator"
FT      5583..5605
FT      /tag= r
FT      stem_loop
FT      WO9408023-A.
FT      14-APR-1994.
FT      01-OCT-1993; 93WO-EP02688.
FT      02-OCT-1992; 92CH-0003124.
FT      15-JUL-1993; 93CH-0002134.
FT      (LONZ ) LONZA AG.
FT      Birch O, Brass J, Fuhrmann M, Shaw N;
FT      WPI; 1994-135587/16.
FT      P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
FT      Biotechnological biotin prodn. using enterobacterial biotin-gene
FT      - providing vitamin H in high yield
FT      Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
FT      The sequence is derived from plasmid pB030A-15/9 contg. the
FT      bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
FT      of biotin, arranged in a transcription unit. Microorganisms
FT      contg. these DNA fragments or plasmids may be used in the prodn.
FT      of biotin. Biotin (Vitamin H) may prevent seborrhea, dermatitis,
FT      CC loss of appetite and tiredness.
FT      SQ Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;
```

Query Match 100.0%; Score 25; DB 15; Length 5872;  
 Best Local Similarity 100.0%; Pred. No. 0.014;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 attgtcgcaagtcacagaattattt 25  
 |||||  
 Db 140 attgtcgcaagtcacagaattattt 164

RESULT 5  
 AAZ56756/C  
 ID AAZ56756 standard; cDNA; 902 BP.

XX AAZ56756;

AC AAZ56756;

XX 23-MAR-2000 (first entry)

XX Human transmembrane protein HTPMN-59 encoding cDNA.  
 DE Human; transmembrane protein; HTPMN; diagnosis; immunospecific;  
 KW antiproliferative; neuroprotective; immune disorder;  
 KW reproductive disorder; smooth muscle disorder; neurological disorder;  
 KW gastrointestinal disorder; developmental disorder;  
 KW cell proliferative disorder; ss.

Wed Oct 10 07:44:26 2001

us-09-396-196f-2.std.rng

```

XX OS Homo sapiens.
XX PN WO9961471-A2.
XX XX
XX PD 02-DEC-1999.
XX XX
XX PF 28-MAY-1999; 99WO-US11904.
XX XX
XX PR 29-MAY-1998; 98US-0087260.
XX PR 02-JUL-1998; 98US-0091674.
XX PR 02-OCT-1998; 98US-0102954.
XX PR 24-NOV-1998; 98US-0109869.
XX XX
XX PA (INCY-) INCYTE PHARM INC.
XX PI Tang YT, Lal P, Hillman JL, Yue H, Guegler KJ, Corley NC;
XX PI Bandman O, Patterson C, Gorgone GA, Kaser MR, Baughn MR;
XX PI Au-Young J;
XX XX
XX DR WPI; 2000-072605/06.
XX DR P-PSDB; AAY57935.
XX XX
XX PT Proteins, polynucleotides, vectors, host cells and antibodies used to
XX PT diagnose, treat or prevent immune, reproductive, smooth muscle,
XX PT neurological, gastrointestinal, developmental and cell proliferative
XX PT disorders -
XX XX
XX PS Claim 9; Page 215; 229pp; English.
XX XX
XX CC AA256698 to AA256776 encode AAY57877 to AAY57955 which represent human
XX CC transmembrane proteins designated HTMPN-1 to HTMPN-79, respectively.
XX CC The transmembrane protein have immunospecific, antiproliferative and
XX CC neuroprotective activities. The human transmembrane proteins,
XX CC polynucleotides encoding them and other compositions and methods from
XX CC the present invention, can be used for the diagnosis, treatment or
XX CC prevention of immune, reproductive, smooth muscle, neurological,
XX CC gastrointestinal, developmental and cell proliferative disorders. The
XX CC HTMPN's can be used to treat or prevent disorders associated with a
XX CC decreased expression or activity of HTMPN.
XX XX
XX SQ Sequence 902 BP; 261 A; 212 C; 196 G; 233 T; 0 other;

Query Match 74.4%; Score 18.6; DB 21; Length 902;
Best Local Similarity 84.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattatt 25
   |||||
Db 434 AATGGCGCTAGTCACAGAAATATAT 410

RESULT 6
AAQ51236
ID AAQ51236 standard; cDNA; 2114 BP.
XX XX
XX AC AAQ51236,
XX XX
XX DT 11-MAY-1994 (first entry)
XX XX
XX DE Plant NADPH cytochrome P450 reductase (ara B).
XX XX
XX KW NADPH cytochrome P450 reductase; functional complementation;
XX KW identification; ss.
XX XX
XX OS Arabidopsis thaliana.
XX XX
XX FH Key Location/Qualifiers
XX FT CDS 36..2114
XX FT /*tag= a
XX FT /product= NADPH cytochrome P450 reductase.
XX FT misc_difference 510..512

/*tag= c
/transl_except= AGA encodes Lys.
misc_difference 543..545
/*tag= d
/transl_except= CGG encodes Lys.
663..665
/*tag= e
/transl_except= CGT encodes Lys.
819..821
/*tag= f
/transl_except= CGG encodes Lys.
915..917
/*tag= g
/transl_except= AGA encodes Lys.
963..965
/*tag= h
/transl_except= CGG encodes Lys.
996..998
/*tag= i
/transl_except= AGG encodes Lys.
1203..1205
/*tag= j
/transl_except= AGA encodes Lys.
1230..1232
/*tag= k
/transl_except= CGA encodes Lys.
1431..1433
/*tag= l
/transl_except= CGT encodes Lys.
1443..1445
/*tag= m
/transl_except= CGT encodes Lys.
1484..1486
/*tag= n
/transl_except= AGA encodes Lys.
1533..1535
/*tag= o
/transl_except= AGA encodes Lys.
1623..1625
/*tag= p
/transl_except= CGA encodes Lys.
1704..1706
/*tag= q
/transl_except= AGA encodes Lys.
1782..1784
/*tag= p
/transl_except= AGA encodes Lys.
1788..1790
/*tag= q
/transl_except= CGA encodes Lys.
1872..1874
/*tag= r
/transl_except= CGT encodes Lys.
1992..1994
/*tag= s
/transl_except= AGG encodes Lys.
2004..2006
/*tag= t
/transl_except= CGA encodes Lys.
2091..2093
/*tag= u
/transl_except= AGA encodes Lys.
2100..2102
/*tag= v
/transl_except= AGA encodes Lys.

WO9321326-A.
28-OCT-1993.
13-APR-1993; 93WO-FR00367.
13-APR-1992; 92FR-0004491.

```

XX PA (ORSA-) ORSAN.  
 XX PI Kaznaier M, Lacroute F, Mignotte-vieux C, Minet M;  
 XX PI Pompon D;  
 XX DR WPI; 1993-351736/44.  
 XX DR P-PSDB; AAR43581.  
 XX PT New DNA encoding plant NADPH cytochrome P450 reductase - cloned  
 XX PT by functional complementation in yeast, also recombinant enzyme  
 XX PT useful in P450 mediated bioconversion processes  
 XX PS  
 XX PS Claim 11; Figure 9; 79pp; French.  
 XX CC A new method for determining whether a DNA sequence encodes an NADPH  
 XX CC cytochrome P450 reductase involves transforming yeasts with plasmids  
 XX CC of a total cDNA bank of plant(s). The yeasts used in the procedure  
 XX CC are incapable of producing their own NADPH cytochrome P450  
 XX CC reductase. They are then exposed to a cytochrome P450 inhibitor at a  
 XX CC level which is lethal to the yeast cells but not to cells which,  
 XX CC because of the transformation, now contain an active NADPH  
 XX CC cytochrome P450 reductase. Surviving clones are then isolated and  
 XX CC plasmid DNA extracted. The gene is inserted into the plasmid at a  
 XX CC site which places it under the control of an inducible promoter.  
 XX SS  
 XX SQ Sequence 2114 BP; 603 A; 400 C; 542 G; 569 T; 0 other;

Query Match 74.4%; Score 18.6; DB 14; Length 2114;  
 Best Local Similarity 84.0%; Pred. No. 13;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 1338 attgttcgaagtcagagaagtcctt 1362

RESULT 7  
 AAZ01425/C  
 ID AAZ01425 standard; DNA; 1038602 BP.  
 XX AC AAZ01425;  
 XX DT 07-OCT-1999 (first entry)  
 XX DE Complete genome sequence of Chlamydia trachomatis.  
 XX KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
 XX KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;  
 XX KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 XX KW bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.  
 XX OS Chlamydia trachomatis.  
 XX PN W09928475-A2.  
 XX PD 10-JUN-1999.  
 XX PF 27-NOV-1998; 98WO-IB01939.  
 XX PR 04-NOV-1998; 98US-0107077.  
 XX PR 28-NOV-1997; 97FR-0015041.  
 XX PR 17-DEC-1997; 97FR-0016034.  
 XX PA (GEST ) GENSET.  
 XX PI Griffais R;  
 XX DR WPI; 1999-371125/31.  
 XX PT Genome sequence of Chlamydia trachomatis  
 XX SS

PS Claim 1; Page 373-656; 1755pp; English.  
 XX CC The present sequence represents the complete genome of Chlamydia  
 XX CC trachomatis. Open reading frames (ORFs) of the genome encode  
 XX CC polypeptides AAY36754-Y37949. The polypeptides can be used as vaccines  
 XX CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also  
 XX CC be used to control growth of the microorganism. Chlamydia trachomatis is  
 XX CC responsible for a large number of diseases, e.g. eye diseases such as  
 XX CC conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion  
 XX CC conjunctivitis; genital diseases such as nongonococcal urethritis;  
 XX CC epididymitis, cervicitis, salpingitis, perihhepatitis, bartholinitis;  
 XX CC pneumopathy in breast feeding infants; and venereal  
 XX CC lymphogranulomatosis. The polypeptides of the invention may be of use in  
 XX CC treating these diseases.  
 XX SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

Query Match 71.2%; Score 17.8; DB 20; Length 1038602;  
 Best Local Similarity 90.5%; Pred. No. 67;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 tcgcaagtcacagaattattt 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 988106 TCGGAAGTCGACAGAAATTTT 988086

RESULT 8  
 AAC44922  
 ID AAC44922 standard; DNA; 564 BP.  
 XX AC AAC44922;  
 XX DT 18-OCT-2000 (first entry)  
 XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 44634.  
 XX KW Hybridisation assay; genetic mapping; gene expression control;  
 XX KW protein identification; signal transduction pathway;  
 XX KW metabolic pathway; promoter; termination sequence; ss.  
 XX OS Arabidopsis thaliana.  
 XX PN EP1033405-A2.  
 XX PD 06-SEP-2000.  
 XX PF 25-FEB-2000; 2000EP-0301439.  
 XX PR 25-FEB-1999; 99US-0121825.  
 XX PR 05-MAR-1999; 99US-0123180.  
 XX PR 09-MAR-1999; 99US-0123548.  
 XX PR 23-MAR-1999; 99US-0125788.  
 XX PR 25-MAR-1999; 99US-0126264.  
 XX PR 29-MAR-1999; 99US-0126785.  
 XX PR 01-APR-1999; 99US-0127462.  
 XX PR 06-APR-1999; 99US-0128234.  
 XX PR 08-APR-1999; 99US-0128714.  
 XX PR 16-APR-1999; 99US-0129845.  
 XX PR 19-APR-1999; 99US-0130077.  
 XX PR 21-APR-1999; 99US-0130449.  
 XX PR 23-APR-1999; 99US-0130510.  
 XX PR 23-APR-1999; 99US-0130891.  
 XX PR 28-APR-1999; 99US-0131449.  
 XX PR 30-APR-1999; 99US-0132048.  
 XX PR 30-APR-1999; 99US-0132407.  
 XX PR 04-MAY-1999; 99US-0132484.  
 XX PR 05-MAY-1999; 99US-0132485.  
 XX PR 06-MAY-1999; 99US-0132486.  
 XX PR 06-MAY-1999; 99US-0132487.  
 XX PR 07-MAY-1999; 99US-0132863.  
 XX PR 11-MAY-1999; 99US-0134256.  
 XX PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136382.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.

PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 70.4%; Score 17.6; DB 21; Length 564;  
 Best Local Similarity 83.3%; Pred. No. 32;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattattt 25  
 DB 446 tggtcgcaagtcacagaattattt 469

RESULT 9  
 AAA61269  
 ID AAA61269 standard; DNA; 3881 BP.

AC AAA61269;  
 DT 18-OCT-2000 (first entry)

DE Human secreted protein gene 10 clone HDPGP94.

KW Human; secreted protein; fusion protein; gene therapy;  
 KW protein therapy; diagnosis; tissue; cancer; tumour; AIDS;  
 KW autoimmune disorder; allergy; cardiovascular; viral; bacterial;  
 KW fungal infection; immunosuppressive; ds.

OS Homo sapiens.

XX WO200029422-A1.

XX 25-MAY-2000.

XX 09-NOV-1999; 99WO-US26409.

XX 12-NOV-1998; 98US-0108207.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ni J, Ruben SM, Rosen CA, Ebner R, Florence KA, Young PE;  
 PI Birse CE, Carter KC, Komatsoulis G;

XX WPI; 2000-387729/33.

XX Novel human secreted proteins useful for diagnosing, preventing,  
 PT treating and ameliorating a medical condition e.g. cardiovascular  
 PT disease -

XX Claim 1; Page 233-234; 295pp; English.

XX The present sequence represents a nucleic acid molecule which encodes a  
 CC secreted human protein. The gene number and the clone it was derived  
 CC from are given in the descriptor line.  
 CC The invention relates to 31 novel genes.  
 CC acid sequences: AAA61260-A61293; amino acid sequences AA812301-B12371)  
 CC which are useful for preventing, treating or ameliorating medical  
 CC conditions e.g. by protein or gene therapy. Also pathological conditions  
 CC can be diagnosed by determining the amount of the new polypeptides in a  
 CC sample or by determining the presence of mutations in the new  
 CC polynucleotides. Specific uses are described for each of the 31  
 CC and include products for the diagnosis or treatment of cancer, tumours,  
 CC AIDS, autoimmune disorders, allergy, cardiovascular disorders, viral,  
 CC bacterial and fungal infection. The genes are used to generate fusion  
 CC proteins by linking to the gene a human immunoglobulin portion (AAA61251)

CC for increasing stability of the fused protein as compared to the  
 CC secreted protein only.

XX Sequence 3881 BP; 1043 A; 610 C; 716 G; 1512 T; 0 other;

Query Match 70.4%; Score 17.6; DB 21; Length 3881;  
 Best Local Similarity 83.3%; Pred. No. 41;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattattt 25  
 DB 2645 ttatccaaagtcacagattattt 2668

RESULT 10  
 AAC24233/c  
 ID AAC24233 standard; cDNA; 386 BP.

XX AAC24233;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 28308.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 28308; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.

XX Sequence 386 BP; 121 A; 55 C; 58 G; 144 T; 8 other;

Query Match 68.8%; Score 17.2; DB 21; Length 386;  
 Best Local Similarity 86.4%; Pred. No. 46;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattatt 23





```

FT      /*tag= am
FT      /number= "7"
FT      intron 12993..13449
FT      /*tag= an
FT      exon   13450..13552
FT      /*tag= ao
FT      /number= "8"
XX
XX WO200104315-A2.
XX
XX PD 18-JAN-2001.
XX
XX PF 12-JUL-2000; 2000WO-US19035.
XX
XX PR 12-JUL-1999; 99US-0143364.
XX
XX PA (TEXA ) UNIV TEXAS A & M SYSTEM.
XX PA (CORR ) CORNELL RES FOUND INC.
XX
XX PI Giovannoni J, Tanksley S, Vrebalov J, Padmanabhan V, Ruezinsky D;
XX PI White R;
XX
XX DR WPI; 2001-103084/11.
XX
XX New isolated nucleic acid sequence comprising RIN (ripening-inhibitor)
XX or MC (macroalyn) genes for use in genetic transformation techniques
XX to manipulate a variety of plant characteristics -
XX
XX PS Disclosure; Fig 7; 167pp; English.
XX
XX The present sequence is tomato chromosome 5 harbouring the RIN and MC
XX genes. The invention relates to the RIN (ripening-inhibitor) and MC
XX (macroalyn) genes. The RIN and MC genes are useful in controlling
XX of fruit ripening and quality, control of sepal development and
XX size, control of senescence, control of pathogen infection, control
XX of ethylene response, and DNA markers for assisted breeding. The
XX RIN and MC genes are used in genetic transformation techniques to
XX manipulate a variety of plant characteristics. Hence these genes
XX represent a valuable new tool for the creation of transgenic plants,
XX preferably having one or more added beneficial characteristics.
XX
XX SQ Sequence 13830 BP; 4991 A; 1594 C; 1910 G; 5335 T; 0 other;

Query Match      68.8%; Score 17.2; DB 22; Length 13830;
Best Local Similarity 86.4%; Pred. No. 76;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 gtcgcaagtcacagaattattt 25
DB 4716 gtggaagtcacagaattattt 4737
|||||

RESULT 12
AAQ56307/c
ID AAQ56307 standard; DNA; 17 BP.
XX
XX AC AAQ56307;
XX
XX DT 13-APR-1994 (first entry)
XX
XX DE BioB DNA primer.
XX
XX KW BioA; BioB; promoter; biotin; operon; primer; ss.
XX
XX OS Synthetic.
XX
XX PN JP05219956-A.
XX
XX PD 31-AUG-1993.
XX
XX PF 14-SEP-1992; 92JP-0244792.

```

```

XX 14-SEP-1992; 92JP-0244792.
XX
XX (SHIS ) SHISEIDO CO LTD.
XX (TAKE ) TAKEDA CHEM IND LTD.
XX
XX WPI; 1993-308323/39.
XX
XX DNA sequence of biotin operon - has base sequence of E. coli
XX mutated by base pair(s) compared to wild type
XX
XX Example 1; Fig 8; 11pp; Japanese.
XX
XX A novel DNA sequence comprises the E.coli biotin operon (BO) in which
XX the control region of BO or the region near the bioB initiation
XX codon is mutated by at least one base pair compared to its
XX wild type. Two primers (AAQ56306-Q56307) are described in Example 1.
XX A microorganism belonging to Escherichia genus, transformed by
XX a recombinant plasmid carrying such DNA can be used for the prodn.
XX of biotin-active substances.
XX
XX SQ Sequence 17 BP; 3 A; 4 C; 4 G; 6 T; 0 other;

Query Match      68.0%; Score 17; DB 14; Length 17;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 gtcgcaagtcacagaat 20
DB 17 GTCGCAAGTCACAGAAT 1
|||||

RESULT 13
AAV88744
ID AAV88744 standard; cDNA; 512 BP.
XX
XX AC AAV88744;
XX
XX DT 12-FEB-1999 (first entry)
XX
XX DE EST clone HK26.
XX
XX KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
XX tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
XX chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
XX receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO9845437-A2.
XX
XX PD 15-OCT-1998.
XX
XX PF 10-APR-1998; 98WO-US06956.
XX
XX PR 10-APR-1997; 97US-0837312.
XX
XX PA (GEMY ) GENETICS INST INC.
XX
XX PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
XX PI Racie LA, Spaulding V, Treacy M;
XX
XX DR WPI; 1999-070078/06.
XX
XX New polynucleotides encoding human secreted proteins - derived from
XX e.g. human blood, kidney, foetal lung, placenta, testes, brain,
XX ovary, pituitary, retina and colon cDNA libraries
XX
XX Claim 1; Page 498; 641pp; English.
XX
XX The present sequence represents an expressed sequence tag (EST), and is
XX a polynucleotide of the invention. The polynucleotides of the invention

```

are all secreted EST sequences isolated from a variety of human tissue sources. The EST sequences and proteins encoded by them are predicted to have useful biological activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals, although no supporting data is given. Suggested activities include nutritional activity, immune stimulating or suppressing activity, haematopoiesis regulating activity, tissue growth activity, haemostatic activity/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, cadherin/tumour invasion suppressor activity, tumour inhibition activity. The EST sequences are also stated to be useful for gene therapy.

Sequence 512 BP; 160 A; 104 C; 96 G; 152 T; 0 other;

Query Match 68.0%; Score 17; DB 20; Length 512;  
Best Local Similarity 80.0%; Pred. No. 60;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25  
||||| | ||||| ||||| ||  
Db 82 attgtcgcaagtcacagaattcatt 106

RESULT 14  
AAV52308/C  
ID AAV52308 standard; DNA; 4597 BP.  
XX  
AC AAV52308;  
XX  
DT 23-OCT-1998 (first entry)  
XX  
DE Streptococcus pneumoniae genome fragment SEQ ID NO:175.  
XX  
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
KW computer readable medium; vaccine; pharmaceutical composition; ds.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO9818931-A2.  
XX  
PD 07-MAY-1998.  
XX  
PF 30-OCT-1997; 97WO-US19588.  
XX  
PR 31-OCT-1996; 96US-0029960.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC..  
XX  
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;  
PI Kunsch CA, Rosen CA;  
XX  
XX WPI; 1998-272225/24.  
XX  
DR Computer-readable medium with recorded Streptococcus pneumoniae  
PT polynucleotide sequences - useful in diagnostic kits and assays, and  
PT pharmaceutical compositions and vaccines for Streptococcus  
PT pneumoniae  
XX  
XX Claim 1; Page 1105-1107; 1409pp; English.  
XX  
XX The present invention describes a computer readable medium which has  
XX the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)  
XX recorded on it, or a representative fragment or a sequence at least 95%  
XX identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in  
XX SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from  
XX Streptococcus pneumoniae. The present invention also describes an  
XX isolated nucleic acid molecule encoding a homologue of any of the  
XX fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the  
XX nucleic acid molecule is produced by a process comprising: (a) screening  
XX a genomic DNA library using as a probe a target sequence defined by any  
XX of the sequences in SEQ ID NO:1 to 391, identifying members of the

CC library which contain sequences that hybridise to the target sequence and  
CC isolating the nucleic acid molecules from the members; or (b) isolating  
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid  
CC molecules whose nucleotide sequence is homologous to amplification  
CC primers derived from the fragment of the S. pneumoniae genome to prime  
CC the amplification and isolating the amplified sequences. The computer  
CC readable medium can be used in a computer-based system for identifying  
CC fragments of the S. pneumoniae genome of commercial importance, or  
CC expression modulating fragments of the S. pneumoniae genome. Products  
CC from the present invention can be used in diagnosis kits and assays, and  
CC pharmaceutical compositions and vaccines for S. pneumoniae.  
XX  
SQ Sequence 4597 BP; 1453 A; 681 C; 957 G; 1506 T; 0 other;

Query Match 68.0%; Score 17; DB 19; Length 4597;  
Best Local Similarity 80.0%; Pred. No. 81;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25  
||||| | ||||| ||||| ||  
Db 4474 ATTTTCATTCCTCCATATTATT 4450

RESULT 15  
AAQ81762/C  
ID AAQ81762 standard; DNA; 10095 BP.  
XX  
AC AAQ81762;  
XX  
DT 03-AUG-1995 (first entry)  
XX  
DE Japanese Black Pine chloroplast P5-P6 fragment.  
XX  
KW Japanese Black Pine; Kuro-matsu; chloroplast; wood marker gene; ds.  
XX  
OS Pinus thunbergii.  
XX  
FH Key Location/Qualifiers  
CDS complement (158..1219)  
FT /\*tag= a  
FT /label= psbA  
FT /note= "see AAQ81763"  
FT complement (1750..1783)  
FT /\*tag= b  
FT /number= 2  
FT /note= "trnk, 2nd. exon (see AAQ81764)"  
FT complement (1784..4285)  
FT /\*tag= c  
FT /note= "trnk intron"  
FT complement (4286..4321)  
FT /\*tag= d  
FT /number= 1  
FT /note= "trnk, 1st. exon, (see AAQ81764)"  
FT complement (5110..6642)  
FT /\*tag= e  
FT /label= ORF510  
FT /note= "see AAQ81770"  
FT complement (6846..6917)  
FT /\*tag= f  
FT /label= trnQ  
FT /note= "does not include start or stop codons (see AAQ81765)"  
FT 7272..7451  
FT /\*tag= g  
FT /label= psbK  
FT /note= "see AAQ81766"  
FT 7891..8049  
FT /\*tag= h  
FT /label= psbI  
FT /note= "see AAQ81767"  
FT complement (8135..8222)  
FT /\*tag= i

FT /label= trns  
 FT /note= "does not include start or stop codons  
 FT (see AAQ81768)"  
 FT 9091..9113  
 FT /\*tag= j  
 FT /number= 1  
 FT /note= "trng, 1st.exon (see AAQ81769)"  
 FT 9114..9853  
 FT /\*tag= k  
 FT /note= "trng intron"  
 FT 9854..9902  
 FT /\*tag= l  
 FT /number= 2  
 FT /note= "trng, 2nd. exon (see AAQ81769)"  
 FT  
 PN JP05276957-A.  
 XX  
 XX 26-OCT-1993.  
 PD  
 XX 13-MAR-1991; 91JP-0072148.  
 PF  
 XX 13-MAR-1991; 91JP-0072148.  
 PR  
 XX (MTS-) MITSUI GYOSAI SHOKUBUTSU BIO KENKYUSHO.  
 XX  
 XX WPI; 1993-373586/47.  
 DR P-PSDB; AAR68890-R68893.  
 XX  
 XX Base sequence contg. chloroplast gene - useful as marker in  
 PT genetic research of chloroplast for developing useful wood  
 PT materials  
 PT  
 XX Claim 1; Fig 1-6; 20pp; Japanese.  
 XX  
 CC This is the fragment of the Japanese Black pine chloroplast genome  
 CC spanning regions P5 and P6. The coding regions it contains (i.e.  
 CC psbA, trnK, trnQ, psbK, psbI, trnS, trnG and ORF510) are each  
 CC claimed and are useful as markers for development of wood  
 CC materials.  
 XX  
 SQ Sequence 10095 BP; 3255 A; 1924 C; 1877 G; 3039 T; 0 other;

Query Match 68.0%; Score 17; DB 14; Length 10095;  
 Best Local Similarity 80.0%; Pred. No. 90;  
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 attgtcgcaagtcacagaattattt 25  
 Db 7846 ATTGTTACAAATCAGCAAGAAATTATT 7822

RESULT 16  
 AAQ11851/c  
 ID AAQ11851 standard; DNA; 3083 BP.  
 XX  
 AC AAQ11851;  
 XX  
 XX 31-JUL-1991 (first entry)  
 DT  
 XX  
 DE Glutamate receptor 3.  
 XX  
 KW Glutamate receptor 3; probe; ligand; drug screening; ss.  
 XX  
 OS Rattus rattus.  
 XX  
 EH Key Location/Qualifiers  
 FT CDS 167..2830  
 FT /\*tag= a  
 FT /product= GR3  
 FT sig\_peptide 167..232  
 FT /\*tag= b  
 FT mat\_peptide 233..2830

FT /\*tag= c  
 XX WO9106648-A.  
 XX  
 XX 16-MAY-1991.  
 PD  
 XX 25-OCT-1990; 90WO-US06153.  
 PF  
 XX 27-OCT-1989; 89US-0428116.  
 PR  
 XX (SALK ) SALK INST FOR BIOL. STUD.  
 XX  
 XX Heinemann SF, Boulter JR, Hollmann M, Bettler B, Jensen JE;  
 PI P-PSDB; AAR11991.  
 DR WPI; 1991-164197/22.  
 DR  
 XX Glutamate receptors - used to screen for functional ligands and  
 PT identify and isolate further receptors  
 PT  
 XX Disclosure; Fig 4; 109pp; English.  
 PS  
 XX GluR3 cDNA was isolated from a rat forebrain cDNA using a  
 CC low-stringency screening protocol and a radiolabelled fragment of the  
 CC GluR1 cDNA as probe. The cDNA is deposited (ATCC 68133).  
 CC The gene and protein can be used in drug screening, to  
 CC determine whether a substance is a functional ligand for the  
 CC receptor by monitoring ion channel activity.  
 CC See also AAQ11849-855.  
 XX  
 SQ Sequence 3083 BP; 877 A; 679 C; 735 G; 792 T; 0 other;  
 Query Match 67.2%; Score 16.8; DB 12; Length 3083;  
 Best Local Similarity 90.0%; Pred. No. 95;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 attgtcgcaagtcacagaatt 20  
 Db 3033 ATTGTCGCAAGTCTCAGAGT 3014  
 RESULT 17  
 AAT21899/c  
 ID AAT21899 standard; cDNA to mRNA; 274 BP.  
 XX  
 AC AAT21899;  
 XX  
 DT 14-AUG-1996 (first entry)  
 DT  
 XX  
 DE Human gene signature HUMGS03441.  
 XX  
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
 KW human; cloning; mapping; non-biased library; diagnosis; detection;  
 KW cell typing; abnormal cell function; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9514772-A1.  
 PD  
 XX 01-JUN-1995.  
 PF  
 XX 11-NOV-1994; 94WO-JP01916.  
 XX  
 PR 12-NOV-1993; 93JP-0355504.  
 XX  
 XX (MATS/) MATSUBARA K.  
 XX (OKUB/) OKUBO K.  
 XX Matsubara K, Okubo K;  
 XX WPI; 1995-206931/27.  
 XX



Query Match 66.4%; Score 16.6; DB 10; Length 1041;  
 Best Local Similarity 82.6%; Pred. No. 1e+02;  
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttgtcgaagtcacagaattatt 24  
 ||||| ||||| || ||||| ||  
 Db 25 ttgtcgaagttactgaattgt 47

RESULT 20  
 AAC76029  
 ID AAC76029 standard; cDNA; 2730 BP.  
 XX  
 AC AAC76029;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORFX ORF1584 polynucleotide sequence SEQ ID NO:3167.  
 XX  
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200058473-A2.  
 PN  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 31-MAR-2000; 2000WO-US08621.  
 XX  
 PR 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 XX Shinkets RA, Leach M;  
 PI  
 XX WPI: 2000-602362/57.  
 DR  
 DR P-PSDB; AAB41820.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 PS Claim 5; Page 2384-2386; 5507pp; English.  
 XX  
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 CC osteopathic; antiparkinsonian; nootropic; neuroprotective;  
 CC antipsoriatic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,

CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 SQ Sequence 2730 BP; 839 A; 505 C; 558 G; 826 T; 2 other;

Query Match 66.4%; Score 16.6; DB 21; Length 2730;  
 Best Local Similarity 82.6%; Pred. No. 1.2e+02;  
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 attgtcgaagtcacagaattatt 23  
 ||| | |||| ||||| ||  
 Db 1292 attatgcaaatcacagaattat 1314

RESULT 21  
 AAF21307  
 ID AAF21307 standard; DNA; 32351 BP.  
 XX  
 AC AAF21307;  
 XX  
 DT 14-MAR-2001 (first entry)  
 XX  
 DE Human low adenosine antisense oligonucleotide related sequence #2874.  
 XX  
 KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;  
 KW human; airway disorder; bronchoconstriction; lung inflammation;  
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;  
 KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;  
 KW respiratory obstruction; pulmonary obstruction; impeded respiration;  
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;  
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;  
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;  
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;  
 KW cancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO2000062736-A2.  
 PN  
 XX  
 PD 26-OCT-2000.  
 XX  
 PF 24-MAR-2000; 2000WO-US08020.  
 XX  
 PR 06-APR-1999; 99US-0127958.  
 XX  
 PA (UYEC-) UNIV EAST CAROLINA.  
 PA (NYCE/) NYCE J W.  
 XX  
 XX Nyce JW;  
 XX  
 XX WPI: 2000-679539/66.  
 DR  
 XX  
 PT Low adenosine (A) content antisense oligonucleotides which do not  
 PT trigger adenosine receptors during metabolism, useful e.g. for treating  
 PT cancers and respiratory obstructions -  
 XX  
 PS Disclosure; Page 1295-1303; 1592pp; English.  
 XX  
 CC The present invention describes low adenosine (A) content antisense  
 CC oligonucleotides and compositions (I) comprising them. In the antisense  
 CC oligonucleotides the A is replaced by a 'Universal' or alternative base.  
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,  
 CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.  
 CC The antisense oligonucleotides and (I) can be used to down-regulate the  
 CC expression and or activity of target polypeptides associated with  
 CC lung/respiratory disorders and malignancies, such as stimulating and

CC activating peptide factors and transmitters, transcription factors,  
 CC immunoglobulins and antibodies, antibody receptors, cytokines and  
 CC chemokines, endogenously produced specific and non-specific enzymes,  
 CC binding proteins, adhesion molecules and their receptors, cytokine and  
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central  
 CC nervous system (CNS) and peripheral nervous and non-nervous system  
 CC receptors, CNS and peripheral nervous and non-nervous system peptide  
 CC transmitters, defensins, growth factors, vasoactive peptides and  
 CC receptors, binding proteins and malignancy associated proteins. The  
 CC antisense oligonucleotides may be used in this way to treat disorders  
 CC including respiratory obstruction (especially pulmonary obstruction  
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)  
 CC and/or surfactant hypoproduction which are associated with a disease or  
 CC condition selected from pulmonary vasoconstriction, inflammation,  
 CC allergies, asthma, impaired respiration, respiratory distress syndrome  
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary  
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),  
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,  
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide  
 CC fragments and antisense oligonucleotides used in the exemplification of  
 CC the present invention.

XX SQ Sequence 32351 BP; 8594 A; 7026 C; 7405 G; 9326 T; 0 other;

Query Match 66.4%; Score 16.6; DB 21; Length 32351;

Best Local Similarity 82.6%; Pred. No. 1.6e+02; Mismatches 0; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattatt 24  
 ||| | ||||| ||||| |  
 Db 10689 ttgccaaagtcacagaattagt 10711

RESULT 22

AAA35185  
 ID AAA35185 standard; DNA; 32351 BP.

XX AC AAA35185;

XX DT 28-JUL-2000 (first entry)

XX DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:59.

XX KW Human: adenosine receptor; low adenosine antisense oligonucleotide;  
 KW phosphorothioate; impaired respiration; inflammation; allergy;  
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;  
 KW antiallergic; antiasthmatic; cytosolic; analgesic; impaired airway;  
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;  
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;  
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;  
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

XX OS Homo sapiens.

XX PN WO200009525-A2.

XX XX 24-FEB-2000.

XX PF 03-AUG-1999; 99WO-US17712.

XX XX 03-AUG-1998; 98US-0095212.

XX XX (UYEC-) UNIV EAST CAROLINA.

XX PI Nyce JW;

XX DR WPI; 2000-205971/18.

XX PT New antisense oligonucleotides useful for treating e.g. pulmonary  
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,  
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or  
 PT cancers

XX PS Disclosure; Page 1212-1219; 1343pp; English.

XX CC The present invention describes a new composition comprising an  
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which  
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or  
 CC inflammation. The ON can have antiinflammatory, antiallergic,  
 CC antiasthmatic, cytosolic and analgesic activities. The compositions are  
 CC useful for the treatment of diseases associated with inflammation,  
 CC impaired airways, including lung disease and diseases whose secondary  
 CC effects afflict the lungs of a subject. They can be used for treating  
 CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,  
 CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic  
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive  
 CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,  
 CC carcinomas, and cancers which may metastasize to the lungs, including  
 CC breast and prostate cancer. The reduction of the adenosine content of  
 CC the ONs reduces side effects. The A-containing ONs break down with the  
 CC release of deoxyadenosine which activates adenosine receptors causing  
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the  
 CC nucleotide sequences given in the sequence listing from the present  
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last  
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences  
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680  
 CC (AAA32323 to AAA33992) are specifically claimed ONs from the present  
 CC invention. N.B. Sequences given in the disclosure of the present  
 CC invention do not match up with their corresponding SEQ ID NO: sequences  
 CC given in the sequence listing.

XX SQ Sequence 32351 BP; 8594 A; 7027 C; 7405 G; 9325 T; 0 other;

Query Match 66.4%; Score 16.6; DB 21; Length 32351;

Best Local Similarity 82.6%; Pred. No. 1.6e+02; Mismatches 0; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattatt 24  
 ||| | ||||| ||||| |  
 Db 10689 ttgccaaagtcacagaattagt 10711

RESULT 23

AAF21311  
 ID AAF21311 standard; DNA; 40298 BP.

XX AC AAF21311;

XX DT 14-MAR-2001 (first entry)

XX DE Human low adenosine antisense oligonucleotide related sequence #2878.

XX KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;  
 KW human; airway disorder; bronchoconstriction; lung inflammation;  
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;  
 KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytosolic;  
 KW respiratory obstruction; pulmonary obstruction; impaired respiration;  
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;  
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;  
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;  
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;  
 KW cancer; ss.

XX OS Homo sapiens.

XX PN WO200062736-A2.

XX XX 26-OCT-2000.

XX PF 24-MAR-2000; 2000WO-US08020.

XX PR 06-APR-1999; 99US-0127958.

XX PA (UYEC-) UNIV EAST CAROLINA.

PA (NYCE/) NYCE J W.  
 XX NYce JW;  
 PI WPI; 2000-679539/66.  
 XX  
 XX Low adenosine (A) content antisense oligonucleotides which do not  
 PT trigger adenosine receptors during metabolism, useful e.g. for treating  
 PT cancers and respiratory obstructions -  
 XX  
 XX Disclosure; Page 1305-1315; 1592pp; English.  
 PS  
 XX The present invention describes low adenosine (A) content antisense  
 CC oligonucleotides and compositions (I) comprising them. In the antisense  
 CC oligonucleotides the A is replaced by a 'universal' or alternative base.  
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,  
 CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.  
 CC The antisense oligonucleotides and (I) can be used to down-regulate the  
 CC expression and/or activity of target polypeptides associated with  
 CC lung/respiratory disorders and malignancies, such as stimulating and  
 CC activating peptide factors and transmitters, transcription factors,  
 CC immunoglobulins and antibodies, antibody receptors, cytokines and  
 CC chemokines, endogenously produced specific and non-specific enzymes,  
 CC binding proteins, adhesion molecules and their receptors, cytokine and  
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central  
 CC nervous system (CNS) and peripheral nervous and non-nervous system  
 CC receptors, CNS and peripheral nervous and non-nervous system peptide  
 CC transmitters, defensins, growth factors, vasoactive peptides and  
 CC receptors, binding proteins and malignancy associated proteins. The  
 CC antisense oligonucleotides may be used in this way to treat disorders  
 CC including respiratory obstruction (especially pulmonary obstruction  
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)  
 CC and/or surfactant hypoproduction which are associated with a disease or  
 CC condition selected from pulmonary vasoconstriction, inflammation,  
 CC allergies, asthma, impaired respiration, respiratory distress syndrome  
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary  
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),  
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,  
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide  
 CC fragments and antisense oligonucleotides used in the exemplification of  
 CC the present invention.  
 XX  
 SQ Sequence 40298 BP; 10485 A; 9119 C; 9484 G; 11210 T; 0 other;  
 Query Match 66.4%; Score 16.6; DB 21; Length 40298;  
 Best Local Similarity 82.6%; Pred. No. 1.7e+02;  
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 ttgtcgcaagtcacagaattatt 24  
 ||| | ||||| ||||| |  
 Db 11514 ttgccaaagtcacagaattagt 11536  
 RESULT 24  
 AAA35189  
 ID AAA35189 standard; DNA; 40298 BP.  
 XX  
 AC AAA35189;  
 XX  
 XX 28-JUL-2000 (first entry)  
 DE  
 XX Human adenosine receptor related polynucleotide 2nd SEQ ID NO:63.  
 XX  
 KW Human; adenosine receptor; low adenosine antisense oligonucleotide;  
 KW phosphorothioate; impaired respiration; inflammation; allergy;  
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;  
 KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;  
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;  
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;  
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;  
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.  
 XX

OS Homo sapiens.  
 XX WO200009525-A2.  
 XX 24-FEB-2000.  
 XX  
 XX 03-AUG-1999; 99WO-US17712.  
 XX  
 PR 03-AUG-1998; 98US-0095212.  
 XX  
 PA (UYEC-) UNIV EAST CAROLINA.  
 XX  
 XX NYce JW;  
 PI WPI; 2000-205971/18.  
 XX  
 XX New antisense oligonucleotides useful for treating e.g. pulmonary  
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,  
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or  
 PT cancers -  
 XX  
 XX Disclosure; Page 1221-1231; 1343pp; English.  
 PS  
 XX The present invention describes a new composition comprising an  
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which  
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or  
 CC inflammation. The ON can have antiinflammatory, antiallergic,  
 CC antiasthmatic, cytostatic and analgesic activities. The compositions are  
 CC useful for the treatment of diseases associated with inflammation,  
 CC impaired airways, including lung disease and diseases whose secondary  
 CC effects afflict the lungs of a subject. They can be used for treating  
 CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,  
 CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic  
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive  
 CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,  
 CC carcinomas, and cancers which may metastasise to the lungs, including  
 CC breast and prostate cancer. The reduction of the adenosine content of  
 CC the ONs reduces side effects. The A-containing ONs break down with the  
 CC release of deoxyadenosine which activates adenosine receptors causing  
 CC bronchoconstriction and inflammation. AAA32313 to AAA35132 represent the  
 CC nucleotide sequences given in the sequence listing from the present  
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last  
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences  
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680  
 CC (AAA32323 to AAA33992) are specifically claimed ONs from the present  
 CC invention. N.B. Sequences given in the disclosure of the present  
 CC invention do not match up with their corresponding SEQ ID NO: sequences  
 CC given in the sequence listing.  
 XX  
 SQ Sequence 40298 BP; 10485 A; 9121 C; 9482 G; 11210 T; 0 other;  
 Query Match 66.4%; Score 16.6; DB 21; Length 40298;  
 Best Local Similarity 82.6%; Pred. No. 1.7e+02;  
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 ttgtcgcaagtcacagaattatt 24  
 ||| | ||||| ||||| |  
 Db 11514 ttgccaaagtcacagaattagt 11536  
 RESULT 25  
 AAT43451  
 ID AAT43451 standard; DNA; 500 BP.  
 XX  
 AC AAT43451;  
 XX  
 XX 08-AUG-1997 (first entry)  
 XX  
 DE ATM gene exon 11.  
 XX  
 KW ATM gene; ataxia-telangiectasia; progressive genetic disorder; antibody;  
 KW central nervous system; immune system; chromosomal instability; therapy;

KW cancer predisposition; radiation sensitivity; cell cycle abnormality;  
 KW multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;  
 KW general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;  
 KW blood vessel; bulbar conjunctiva; facial skin; A-T; ss.  
 XX

OS Homo sapiens.

XX Key Location/Qualifiers

FT intron 1..110

FT /\*tag= a

FT /number= 10

FT 111..270

FT /\*tag= b

FT /number= 11

FT 271..500

FT /\*tag= c

FT /number= 11

PN W09636691-A1.

XX 21-NOV-1996.

PF 16-MAY-1996; 96WO-US07025.

PR 08-APR-1996; 96US-0629001.

PR 16-MAY-1995; 95US-0441822.

XX (KOHN/) KOHN K I.

PA (UTRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.

XX Shiloh Y;

XX WPI; 1997-012070/01.

XX New isolated ataxia-telangiectasia gene - used to develop prods. for  
 the study, diagnosis and treatment of ataxia-telangiectasia.

PS Claim 1; Page 59; 153pp; English.

CC AAT43444-T43496 represent exons of the ATM gene of the invention.  
 CC Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or  
 CC deletions in the coding region of the ATM gene. A-T is a progressive  
 CC genetic disorder affecting the central nervous and immune systems. A-T  
 CC involves chromosomal instability, cancer predisposition, radiation  
 CC sensitivity, and cell cycle abnormalities. A-T is a multi-system disease  
 CC inherited in an autosomal recessive manner. Cerebellar ataxia that  
 CC gradually develops into general motor dysfunction is the first clinical  
 CC hallmark of A-T, and results from progressive loss of Purkinje cells in  
 CC the cerebellum. Oculocutaneous telangiectasia (dilation of blood  
 CC vessels) develops in the bulbar conjunctiva and facial skin, and is  
 CC later accompanied by graying of the hair and atrophic changes in the  
 CC skin. The co-occurrence of cerebellar ataxia and telangiectases in the  
 CC conjunctivae and occasionally on the facial skin (the second early  
 CC hallmark of A-T) usually establishes the differential diagnosis of A-T  
 CC from other cerebellar ataxias. The ATM gene can be used in methods for  
 CC detecting carriers of a defective gene that causes A-T. The gene can also  
 CC be used to generate antibodies. The antibodies and methods can be used in  
 CC the study, diagnosis and therapy of A-T.

XX Sequence 500 BP; 164 A; 93 C; 90 G; 153 T; 0 other;

Query Match 64.8%; Score 16.2; DB 18; Length 500;

Best Local Similarity 85.7%; Pred. No. 1.4e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 tcgcaagtcacagaattattt 25

II IIIIIIIIIIIIIIIIIII

Db 245 tcagaagtcacagaatgattt 265

RESULT 26

AAC57499/c

ID AAC57499 standard; DNA; 1001 BP.

XX AAC57499;

XX 25-JAN-2001 (first entry)

XX Arachidonic acid metabolism related genomic biallelic marker #133.

XX Human; biallelic marker; arachidonic acid metabolism; genotyping;  
 KW detection; hybridisation; phenotype; haplotype; SNP; polymorphic base;  
 KW single nucleotide polymorphism; hybridisation assay; sequencing assay;  
 KW specific amplification assay; identification; ERBM; 12-LO-RBM;  
 KW eicosanoid-related biallelic marker; 12-LO-related biallelic marker; ds.

XX Homo sapiens.

XX W0200047771-A2.

XX 17-AUG-2000.

XX 11-FEB-2000; 2000WO-IB00184.

XX 12-FEB-1999; 99US-0119917.

XX 23-MAR-1999; 99US-0275267.

XX 07-MAY-1999; 99US-0133200.

XX (GEST ) GENSET.

XX Blumenfeld M, Bougueleret L, Chumakov I;

XX WPI; 2000-571881/53.

XX Novel biallelic markers useful for detecting conditions and genotypes  
 associated with arachidonic acid metabolism -

PS Claim 13; Page 343; 802pp; English.

CC The present invention describes polynucleotides including biallelic  
 CC markers derived from genes involved in arachidonic acid metabolism and  
 CC from genomic regions flanking those genes. Methods from the present  
 CC invention may be used to select individuals for clinical trials and  
 CC predict responses to treatment with drugs. The polynucleotides may be  
 CC used in hybridisation assays, sequencing assays and specific  
 CC amplification assays for identifying an eicosanoid-related biallelic  
 CC marker (ERBM) or 12-LO-related biallelic marker, and for amplifying a  
 CC segment of nucleotides containing an ERBM. The polynucleotides are  
 CC useful in diagnostic kits. The markers may be used to detect conditions  
 CC and genotypes associated with arachidonic acid metabolism. AAC57367 to  
 CC AAC58018 and AAB4019 and AAB4020 represent sequences used in the  
 CC exemplification of the present invention.

CC N.B. Polymorphic bases (single nucleotide polymorphisms also known as  
 CC SNPs) in the polynucleotide sequences from the present invention have  
 CC been given as their corresponding degenerate bases e.g. a polymorphic  
 CC base of C or T has been given as Y.

XX Sequence 1001 BP; 328 A; 186 C; 179 G; 307 T; 1 other;

Query Match 64.8%; Score 16.2; DB 21; Length 1001;

Best Local Similarity 85.7%; Pred. No. 1.6e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 tcgcaagtcacagaattattt 25

II IIIIIIIIIIIIIIIIIII

Db 373 TCCCAATCACTGAATTATT 353

RESULT 27

AAC57500/c

ID AAC57500 standard; DNA; 1001 BP.

XX AAC57500;





OS Zea mays.  
 PN W09413825-A.  
 XX  
 PD 23-JUN-1994.  
 XX  
 PF 14-DEC-1993; 93WO-US12146.  
 XX  
 PR 15-DEC-1992; 92US-0995658.  
 XX  
 PA (PION-) PIONEER HI-BRED INT INC.  
 XX  
 PI Briggs SP, Johal GS;  
 XX  
 DR WPI; 1994-217898/26.  
 XX  
 PT Hml gene conferring race-resistance to Cochliobolus carbonum to  
 PT maize - for use as a selectable marker for transformed maize  
 PT cells  
 XX  
 PS Disclosure; Page 18; 19pp; English.  
 XX  
 CC Transposon mutagenesis was used to tag, clone and characterize the  
 CC maize Hml gene. Genomic and cDNA sequences of the Hml gene are  
 CC provided in AAQ68433 and AAQ68434, respectively.  
 XX  
 SQ Sequence 1374 BP; 285 A; 406 C; 432 G; 251 T; 0 other;

Query Match 64.8%; Score 16.2; DB 15; Length 1374;  
 Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaatt 21  
 ||| ||||| ||||| |||||  
 DB 1327 ATTATCGCAAGTCATAAAATT 1307

RESULT 30  
 ID AAQ99463/c  
 AC AAQ99463  
 XX  
 DT 30-DEC-1995 (first entry)  
 XX  
 DE Maize Hml gene cDNA.

KW Hml gene; Cochliobolus carbonum Nelson race 1; fungus;  
 KW fungal disease toxin; resistance; ss.  
 XX  
 OS Zea mays.  
 XX  
 PN W09507989-A.  
 XX  
 PD 23-MAR-1995.  
 XX  
 PF 19-SEP-1994; 94WO-US10497.  
 XX  
 PR 17-SEP-1993; 93US-0123761.  
 XX  
 PA (PION-) PIONEER HI-BRED INT INC.  
 XX  
 PI Briggs SP, Johal GS;  
 XX  
 DR WPI; 1995-131357/17.  
 XX  
 CC Disease resistance gene Hml from maize - is used to confer  
 PT resistance to Cochliobolus carbonum.  
 XX  
 PS Claim 2; Page 23; 26pp; English.  
 XX  
 CC A DNA sequence is claimed having 90% translational homology to

CC genomic or cDNA sequence of maize Hml gene (AAQ99462 & AAQ99463). The  
 CC gene Hml controls resistance to Cochliobolus carbonum Nelson race 1.  
 CC The Hml gene can be used in conjunction with HC-toxin in a  
 CC selectable marker system for use in maize transformation.  
 CC This gene can also be inserted into the genomes of maize varieties  
 CC which lack the gene to produce transformants that are resistance  
 CC to disease caused by C. carbonum. The Hml resistance gene is one of  
 CC a family of homologous disease resistance genes in maize and other  
 CC crops. The invention encompasses methods for identification of  
 CC homologous disease resistance genes by using this gene or a  
 CC fragment of it as a probe.  
 XX  
 SQ Sequence 1374 BP; 287 A; 405 C; 432 G; 250 T; 0 other;

Query Match 64.8%; Score 16.2; DB 16; Length 1374;  
 Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaatt 21  
 ||| ||||| ||||| |||||  
 DB 1327 ATTATCGCAAGTCATAAAATT 1307

RESULT 31  
 ID AAQ30602  
 XX AAQ30602 standard; DNA; 1771 BP.

AC AAQ30602;

DT 18-JAN-2000 (first entry)

DE Mouse integrin alpha 4 coding sequence.

XX Human; integrin; antisense; oligonucleotide; inhibition; expression;  
 KW very late antigen; CD49d; CD29; cell surface; leucocyte; adhesion;  
 KW vascular endothelial cell; vascular endothelium; migration; inflammation;  
 KW atherosclerosis; allergy; asthma; rheumatoid arthritis; tumor;  
 KW metastasis; circulatory system; autoimmune disease; Grave's disease;  
 KW Hashimoto's thyroiditis; encephalomyelitis; multiple sclerosis; ds.

OS Mus sp.

PN US5968826-A.

PD 19-OCT-1999.

PF 05-OCT-1998; 98US-0166203.

PR 05-OCT-1998; 98US-0166203.

PA (ISIS-) ISIS PHARM INC.

PI Bennett CF, Cowser LM, Condon TP;

PN WPI; 1999-590416/50.

DR P-PSDB; AAY4154.

XX Antisense inhibition of integrin alpha4 expression useful for treating  
 PT inflammatory diseases such as atherosclerosis, allergies, asthma and  
 PT arthritis -

PS Example 12; Column 71-74; 40pp; English.

CC This sequence represents the coding region of the mouse integrin  
 CC alpha4 gene. The invention relates to the generation of antisense  
 CC oligonucleotides targeted to the integrin alpha4 gene which are used  
 CC for inhibiting expression of the integrin alpha4 mRNA or protein.  
 CC Integrin alpha4 is a component of Very Late Antigen (VLA)-4 (also  
 CC called alpha4beta1 and CD49d/CD29). VLA-4 is expressed on the cell  
 CC surfaces of leucocytes and vascular endothelial cells and mediates the  
 CC adhesion of leucocytes to the vascular endothelium prior to migration  
 CC into the surrounding tissues. This migration is an essential step in

CC inflammation and hence VLA-4 (and consequently integrin alpha4) is a  
 CC potential therapeutic target for treating inflammatory diseases and  
 CC the damaging effects of excessive inflammation. These disorders include  
 CC atherosclerosis, allergies, asthma, rheumatoid arthritis and tumor cell  
 CC metastasis (VLA-4 is involved in migration of the tumor cells through  
 CC the extracellular matrix into the circulatory system). VLA-4 is also  
 CC involved in a number of autoimmune diseases such as Grave's disease,  
 CC Hashimoto's thyroiditis, encephalomyelitis (EAE), multiple sclerosis,  
 CC VLA-4 may also be involved in promoting adhesion (i.e. retention) of  
 CC hemopoietic stem cells in bone marrow and in allograft rejection.

XX Sequence 1771 BP; 371 A; 471 C; 519 G; 410 T; 0 other;

Query Match 64.8%; Score 16.2; DB 20; Length 1771;  
 Best Local Similarity 85.7%; Pred. No. 1.7e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtgcgaagtcacagaatt 21  
 ||| |||||  
 Db 306 agtgaggcaagtcacagaatt 326

RESULT 32  
 AAQ68433/c  
 ID AAQ68433 standard; DNA; 5198 BP.

XX AAQ68433;

DT 26-JAN-1995 (first entry)

DE Maize Hml gene.

KW Hml gene; HC toxin; race-specific resistance; disease-resistance;  
 KW Cochliobolus carbonum; selectable marker; transformation;  
 KW transposon tagging; ds.

OS Zea mays.

PN WO9413825-A.

PD 23-JUN-1994.

PF 14-DEC-1993; 93WO-US12146.

PR 15-DEC-1992; 92US-0995659.

PA (PION-) PIONEER HI-BRED INT INC.

PI Briggs SP, Johal GS;

DR WPI; 1994-217898/26.

XX Hml gene conferring race-resistance to Cochliobolus carbonum to  
 PT maize - for use as a selectable marker for transformed maize  
 PT cells

PS Disclosure; Page 14-16; 19pp; English.

XX Transposon mutagenesis was used to tag, clone and characterize the  
 CC maize Hml gene. Genomic and cDNA sequences of the Hml gene are  
 CC provided in AAQ68433 and AAQ68434, respectively.

XX Sequence 5198 BP; 1376 A; 1160 C; 1135 G; 1527 T; 0 other;

Query Match 64.8%; Score 16.2; DB 15; Length 5198;  
 Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtgcgaagtcacagaatt 21  
 ||| |||||  
 Db 4889 ATTATCGCAAGTCATAAAATT 4869

RESULT 33  
 AAQ99462/c  
 ID AAQ99462 standard; DNA; 5198 BP.

XX AAQ99462;

DT 30-DEC-1995 (first entry)

DE Maize Hml genomic DNA.

XX Hml gene; Cochliobolus carbonum Nelson race 1; fungus;  
 KW fungal disease toxin; resistance; ss.

OS Zea mays.

PN WO9507989-A.

PD 23-MAR-1995.

PF 19-SEP-1994; 94WO-US10497.

PR 17-SEP-1993; 93US-0123761.

PA (PION-) PIONEER HI-BRED INT INC.

PI Briggs SP, Johal GS;

DR WPI; 1995-131357/17.

PT Disease resistance gene Hml from maize - is used to confer  
 resistance to Cochliobolus carbonum.

PS Claim 2; Page 20-22; 26pp; English.

XX A DNA sequence is claimed having 90% translational homology to  
 CC genomic or cDNA sequence of maize Hml gene (AAQ99462 & AAQ99463). The  
 CC gene Hml controls resistance to Cochliobolus carbonum Nelson race 1.  
 CC The Hml gene can be used in conjunction with HC-toxin in a  
 CC selectable marker system for use in maize transformation.

CC This gene can also be inserted into the genomes of maize varieties  
 CC which lack the gene to produce transformants that are resistance  
 CC to disease caused by C. carbonum. The Hml resistance gene is one of  
 CC a family of homologous disease resistance genes in maize and other  
 CC crops. The invention encompasses methods for identification of  
 CC homologous disease resistance genes by using this gene or a  
 CC fragment of it as a probe.

XX Sequence 5198 BP; 1375 A; 1159 C; 1136 G; 1528 T; 0 other;

Query Match 64.8%; Score 16.2; DB 16; Length 5198;  
 Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtgcgaagtcacagaatt 21  
 ||| |||||  
 Db 4889 ATTATCGCAAGTCATAAAATT 4869

RESULT 34  
 AAT68772  
 ID AAT68772 standard; DNA; 8341 BP.

XX AAT68772;

DT 19-AUG-1997 (first entry)

DE ATM mutant 4777del830 coding sequence.

XX ATM gene; ataxia-telangiectasia; progressive genetic disorder; antibody;  
 KW central nervous system; immune system; chromosomal instability; therapy;

KW cancer predisposition; radiation sensitivity; cell cycle abnormality;  
 KW multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;  
 KW general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;  
 KW blood vessel; bulbar conjunctiva; facial skin; A-T; mutein; ss.  
 XX Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 1..4878  
 FT /\*tag= a  
 FT /transl\_except= (pos: 787..789, aa: Val)  
 FT /transl\_except= (pos: 1660..1662, aa: Ala)  
 FT /transl\_except= (pos: 1909..1911, aa: His)  
 FT /transl\_except= (pos: 2248..2250, aa: Asn)  
 FT misc\_difference 4776..4777  
 FT /\*tag= b  
 FT /note= "site of 830 nucleotide deletion"  
 XX WO9636691-A1.  
 XX 21-NOV-1996.  
 XX 16-MAY-1996; 96WO-US07025.  
 XX 08-APR-1996; 96US-0629001.  
 XX 16-MAY-1995; 95US-0441822.  
 XX (KOHN/) KOHN K I.  
 XX (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.  
 XX Shiloh Y;  
 XX WPI; 1997-012070/01.  
 XX P-PSDB; AAW19689.  
 XX New isolated ataxia-telangiectasia gene - used to develop prods. for  
 XX the study, diagnosis and treatment of ataxia-telangiectasia.  
 XX Claim 5; Page -: 153pp; English.  
 XX AAT68731-T68785 represent mutations of the ATM gene of the invention  
 XX (see AAT43497 for wild type sequence). The ATM gene of the invention, is  
 XX located in the human genome at chromosome segment 11q22-23.  
 XX Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or  
 XX deletions (such as represented by this sequence) in the coding region of  
 XX the ATM gene. A-T is a progressive genetic disorder affecting the central  
 XX nervous and immune systems. A-T involves chromosomal instability, cancer  
 XX predisposition, radiation sensitivity, and cell cycle abnormalities. A-T  
 XX is a multi-system disease inherited in an autosomal recessive manner.  
 XX The wild type gene, can be used in methods for detecting carriers of a  
 XX defective gene that causes A-T. The gene can also be used to generate  
 XX antibodies. The methods and antibodies can be used in the study,  
 XX diagnosis and therapy of A-T.  
 XX Sequence 8341 BP; 2707 A; 1495 C; 1752 G; 2387 T; 0 other;  
 Query Match 64.8%; Score 16.2; DB 18; Length 8341;  
 Best Local Similarity 85.7%; Pred. No. 2.le+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 5 tcgcaagtcacagaattatt 25  
 || |||||  
 Db 1200 tcagaagtcacagaatgatt 1220  
 RESULT 35  
 AAT68735  
 ID AAT68735 standard; DNA; 8767 BP.  
 XX  
 AC AAT68735;  
 XX  
 DT 18-AUG-1997 (first entry)

XX ATM mutant 8269del403 coding sequence.  
 XX  
 KW ATM gene; ataxia-telangiectasia; progressive genetic disorder; antibody;  
 KW central nervous system; immune system; chromosomal instability; therapy;  
 KW cancer predisposition; radiation sensitivity; cell cycle abnormality;  
 KW multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;  
 KW general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;  
 KW blood vessel; bulbar conjunctiva; facial skin; A-T; mutein; ss.  
 XX Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 1..8280  
 FT /\*tag= a  
 FT /transl\_except= (pos: 787..789, aa: Val)  
 FT /transl\_except= (pos: 1660..1662, aa: Ala)  
 FT /transl\_except= (pos: 1909..1911, aa: His)  
 FT /transl\_except= (pos: 2248..2250, aa: Asn)  
 FT misc\_difference 8268..8269  
 FT /\*tag= b  
 FT /note= "site of 403 nucleotide deletion"  
 XX WO9636691-A1.  
 XX 21-NOV-1996.  
 XX 16-MAY-1996; 96WO-US07025.  
 XX 08-APR-1996; 96US-0629001.  
 XX 16-MAY-1995; 95US-0441822.  
 XX (KOHN/) KOHN K I.  
 XX (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.  
 XX Shiloh Y;  
 XX WPI; 1997-012070/01.  
 XX P-PSDB; AAW19652.  
 XX New isolated ataxia-telangiectasia gene - used to develop prods. for  
 XX the study, diagnosis and treatment of ataxia-telangiectasia.  
 XX Claim 5; Page -: 153pp; English.  
 XX AAT68731-T68785 represent mutations of the ATM gene of the invention  
 XX (see AAT43497 for wild type sequence). The ATM gene of the invention, is  
 XX located in the human genome at chromosome segment 11q22-23.  
 XX Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or  
 XX deletions (such as represented by this sequence) in the coding region of  
 XX the ATM gene. A-T is a progressive genetic disorder affecting the central  
 XX nervous and immune systems. A-T involves chromosomal instability, cancer  
 XX predisposition, radiation sensitivity, and cell cycle abnormalities. A-T  
 XX is a multi-system disease inherited in an autosomal recessive manner.  
 XX The wild type gene, can be used in methods for detecting carriers of a  
 XX defective gene that causes A-T. The gene can also be used to generate  
 XX antibodies. The methods and antibodies can be used in the study,  
 XX diagnosis and therapy of A-T.  
 XX Sequence 8767 BP; 2868 A; 1569 C; 1828 G; 2502 T; 0 other;  
 Query Match 64.8%; Score 16.2; DB 18; Length 8767;  
 Best Local Similarity 85.7%; Pred. No. 2.le+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 5 tcgcaagtcacagaattatt 25  
 || |||||  
 Db 1200 tcagaagtcacagaatgatt 1220  
 RESULT 36  
 AAT73812

ID AAT73812 standard; cDNA; 8768 BP.  
 AC AAT73812;  
 XX 30-MAR-1998 (first entry)  
 DT  
 XX ATM gene variant 8269del403 coding sequence.  
 DE  
 XX Ataxia-telangiectasia; A-T; mutated; ATM; 11q22-23; signal transduction;  
 KW DNA damage; cell cycle control; screening; gene therapy; catalytic;  
 KW phosphatidylinositol-3 kinase; PI-3; cancer predisposition; carrier; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..8409  
 FT /\*tag= a  
 FT /transl\_except= (pos: 787..789, aa: Val)  
 FT /transl\_except= (pos: 1498..1500, aa: Lys)  
 FT /transl\_except= (pos: 1660..1662, aa: Ala)  
 FT /transl\_except= (pos: 1909..1911, aa: His)  
 FT /transl\_except= (pos: 2248..2250, aa: Asn)  
 FT misc\_difference 8268..8269  
 FT /\*tag= b  
 FT /note= "site of 403 nucleotide deletion"  
 FT  
 XX WO9636695-A1.  
 PN 21-NOV-1996.  
 PD 16-MAY-1996; 96WO-US07040.  
 PF 28-JUL-1995; 95US-0508836.  
 PR 16-MAY-1995; 95US-0441822.  
 PR 21-JUN-1995; 95US-0493092.  
 XX  
 XX (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PA Collins FS, Shiloh Y, Tagle DA;  
 PI WPI; 1997-012074/01.  
 DR  
 XX New gene ATM implicated in ataxia-telangiectasia and related protein  
 PT - useful in screening methods, partic. for identifying disease  
 PT carriers  
 XX  
 PS Example 3; Page -: 127pp; English.  
 XX  
 CC This sequence is a variant of the ATM gene from a patient with classical  
 CC ataxia-telangiectasia (A-T), a progressive genetic disorder affecting  
 CC the central nervous and immune systems. The ATM gene, located at  
 CC chromosome 11q22-23, is probably involved in a novel signal transduction  
 CC system that links DNA damage surveillance to cell cycle control. The  
 CC wild-type ATM gene open reading frame (AAT45419) encodes a protein  
 CC (AAW07655) which has a highly conserved C-terminal region showing high  
 CC sequence homology to the catalytic domain of phosphatidylinositol-3  
 CC kinases. A-T mutations affect a variety of tissues and lead to cancer  
 CC predisposition. Identification of A-T carriers, by analysis at nucleic  
 CC acid or protein levels, allows better supervision and treatment of such  
 CC subjects who are at increased risk of developing cancer and are  
 CC particularly sensitive to radiation. The transgenic animals and  
 CC transformed cells are useful as models of the human disease. Also viral  
 CC vectors expressing the ATM protein can be used in gene therapy of A-T.  
 XX  
 SQ Sequence 8768 BP; 2868 A; 1569 C; 1829 G; 2502 T; 0 other;

Query Match 64.8%; Score 16.2; DB 18; Length 8768;  
 Best Local Similarity 85.7%; Pred. No. 2.1e-02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 5 tcgcaagtcacagaattattt 25

Db 1200 tcagaagtcacagaattt 1220  
 || |||||  
 RESULT 37  
 AAT68758  
 ID AAT68758 standard; DNA; 8799 BP.  
 XX  
 AC AAT68758;  
 XX 19-AUG-1997 (first entry)  
 DT  
 XX ATM mutant 2467del372 coding sequence.  
 DE  
 XX ATM gene; ataxia-telangiectasia; progressive genetic disorder; antibody;  
 KW central nervous system; immune system; chromosomal instability; therapy;  
 KW cancer predisposition; radiation sensitivity; cell cycle abnormality;  
 KW multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;  
 KW general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;  
 KW blood vessel; bulbar conjunctiva; facial skin; A-T; mutein; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..8799  
 FT /\*tag= a  
 FT /transl\_except= (pos: 787..789, aa: Val)  
 FT /transl\_except= (pos: 1660..1662, aa: Ala)  
 FT /transl\_except= (pos: 1909..1911, aa: His)  
 FT /transl\_except= (pos: 2248..2250, aa: Asn)  
 FT misc\_difference 2466..2467  
 FT /\*tag= b  
 FT /note= "site of 174 nucleotide deletion"  
 FT  
 XX WO9636691-A1.  
 PN 21-NOV-1996.  
 PD 16-MAY-1996; 96WO-US07025.  
 PF 08-APR-1996; 96US-0629001.  
 PR 16-MAY-1995; 95US-0441822.  
 XX  
 XX (KOHN/) KOHN K I.  
 PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.  
 PA Shiloh Y;  
 PI WPI; 1997-012070/01.  
 DR P-PSDB; AAW19675.  
 XX  
 CC New isolated ataxia-telangiectasia gene - used to develop prods. for  
 CC the study, diagnosis and treatment of ataxia-telangiectasia.  
 XX  
 PS Claim 5; Page -: 153pp; English.  
 XX  
 CC AAT68731-T68785 represent mutations of the ATM gene of the invention  
 CC (see AAT43497 for wild type sequence). The ATM gene of the invention, is  
 CC located in the human genome at chromosome segment 11q22-23.  
 CC Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or  
 CC deletions (such as represented by this sequence) in the coding region of  
 CC the ATM gene. A-T is a progressive genetic disorder affecting the central  
 CC nervous and immune systems. A-T involves chromosomal instability, cancer  
 CC predisposition, radiation sensitivity, and cell cycle abnormalities. A-T  
 CC is a multi-system disease inherited in an autosomal recessive manner.  
 CC The wild type gene, can be used in methods for detecting carriers of a  
 CC defective gene that causes A-T. The gene can also be used to generate  
 CC antibodies. The methods and antibodies can be used in the study,  
 CC diagnosis and therapy of A-T.  
 XX  
 SQ Sequence 8799 BP; 2870 A; 1565 C; 1842 G; 2522 T; 0 other;

```

Query Match      64.8%; Score 16.2; DB 18; Length 8799;
Best Local Similarity 85.7%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 tcgaagtcacagaattattt 25
   || ||||| ||||| |||||
Db 1200 tcagaagtcacagaattattt 1220

RESULT 38
AAT68750
ID AAT68750 standard; DNA; 8816 BP.
XX
AC AAT68750;
XX
DT 18-AUG-1997 (first entry)
XX
DE ATM mutant 5320del355 coding sequence.
XX
KW ATM gene; ataxia-telangiectasia; progressive genetic disorder; antibody;
KW central nervous system; immune system; chromosomal instability; therapy;
KW cancer predisposition; radiation sensitivity; cell cycle abnormality;
KW multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;
KW general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;
KW blood vessel; bulbar conjunctiva; facial skin; A-T; mutein; ss.
XX
OS Homo sapiens.
FH
FT Key Location/Qualifiers
FT CDS 1..5394
FT /*tag= a
FT /transl_except= (pos: 787..789, aa: Val)
FT /transl_except= (pos: 1660..1662, aa: Ala)
FT /transl_except= (pos: 1909..1911, aa: His)
FT /transl_except= (pos: 2248..2250, aa: Asn)
FT misc_difference 5319..5320
FT /*tag= b
FT /*note= "site of 355 nucleotide deletion"
FT
XX
PN WO9636691-Al.
XX
PD 21-NOV-1996.
XX
PF 16-MAY-1996; 96WO-US07025.
XX
PR 08-APR-1996; 96US-0629001.
XX
PR 16-MAY-1995; 95US-0441822.
XX
PA (KOHN/) KOHN K I.
PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
XX
PI Shiloh Y;
XX
WPI; 1997-012070/01.
XX
P-PSDB; AAW19667.
XX
PT New isolated ataxia-telangiectasia gene - used to develop prods. for
PT the study, diagnosis and treatment of ataxia-telangiectasia.
XX
PS Claim 5; Page -; 153pp; English.
XX
AAT68731-T68785 represent mutations of the ATM gene of the invention
CC (see AAT43497 for wild type sequence). The ATM gene of the invention, is
CC located in the human genome at chromosome segment 11q22-23.
CC Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or
CC deletions (such as represented by this sequence) in the coding region of
CC the ATM gene. A-T is a progressive genetic disorder affecting the central
CC nervous and immune systems. A-T involves chromosomal instability, cancer
CC predisposition, radiation sensitivity, and cell cycle abnormalities. A-T
CC is a multi-system disease inherited in an autosomal recessive manner.
CC The wild type gene, can be used in methods for detecting carriers of a
CC defective gene that causes A-T. The gene can also be used to generate
CC antibodies. The methods and antibodies can be used in the study,

```

```

CC diagnosis and therapy of A-T.
XX
SQ Sequence 8816 BP; 2874 A; 1567 C; 1855 G; 2520 T; 0 other;

Query Match      64.8%; Score 16.2; DB 18; Length 8816;
Best Local Similarity 85.7%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 tcgaagtcacagaattattt 25
   || ||||| ||||| |||||
Db 1200 tcagaagtcacagaattattt 1220

RESULT 39
AAT68740
ID AAT68740 standard; DNA; 8873 BP.
XX
AC AAT68740;
XX
DT 18-AUG-1997 (first entry)
XX
DE ATM mutant 7630del298 coding sequence.
XX
KW ATM gene; ataxia-telangiectasia; progressive genetic disorder; antibody;
KW central nervous system; immune system; chromosomal instability; therapy;
KW cancer predisposition; radiation sensitivity; cell cycle abnormality;
KW multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;
KW general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;
KW blood vessel; bulbar conjunctiva; facial skin; A-T; mutein; ss.
XX
OS Homo sapiens.
FH
FT Key Location/Qualifiers
FT CDS 1..7638
FT /*tag= a
FT /transl_except= (pos: 787..789, aa: Val)
FT /transl_except= (pos: 1660..1662, aa: Ala)
FT /transl_except= (pos: 1909..1911, aa: His)
FT /transl_except= (pos: 2248..2250, aa: Asn)
FT misc_difference 7788..7789
FT /*tag= b
FT /*note= "site of 298 nucleotide deletion"
FT
XX
PN WO9636691-Al.
XX
PD 21-NOV-1996.
XX
PF 16-MAY-1996; 96WO-US07025.
XX
PR 08-APR-1996; 96US-0629001.
XX
PR 16-MAY-1995; 95US-0441822.
XX
PA (KOHN/) KOHN K I.
PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
XX
PI Shiloh Y;
XX
WPI; 1997-012070/01.
XX
P-PSDB; AAW19657.
XX
PT New isolated ataxia-telangiectasia gene - used to develop prods. for
PT the study, diagnosis and treatment of ataxia-telangiectasia.
XX
PS Claim 5; Page -; 153pp; English.
XX
AAT68731-T68785 represent mutations of the ATM gene of the invention
CC (see AAT43497 for wild type sequence). The ATM gene of the invention, is
CC located in the human genome at chromosome segment 11q22-23.
CC Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or
CC deletions (such as represented by this sequence) in the coding region of
CC the ATM gene. A-T is a progressive genetic disorder affecting the central
CC nervous and immune systems. A-T involves chromosomal instability, cancer

```

CC predisposition, radiation sensitivity, and cell cycle abnormalities. A-T  
 CC is a multi-system disease inherited in an autosomal recessive manner.  
 CC The wild type gene, can be used in methods for detecting carriers of a  
 CC defective gene that causes A-T. The gene can also be used to generate  
 CC antibodies. The methods and antibodies can be used in the study,  
 CC diagnosis and therapy of A-T.  
 XX  
 SQ Sequence 8873 BP; 2878 A; 1582 C; 1859 G; 2554 T; 0 other;

Query Match 64.8%; Score 16.2; DB 18; Length 8873;  
 Best Local Similarity 85.7%; Pred. No. 2.1e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 tcgcaagtcacagaattattt 25  
 || ||||| ||||| |||||  
 Db 1200 tcagaagtcacagaattattt 1220

RESULT 40  
 AAT68771  
 ID AAT68771 standard; DNA; 8970 BP.  
 XX  
 AC AAT68771;  
 XX  
 DT 19-AUG-1997 (first entry)  
 XX  
 DE ATM mutant 1407del201 coding sequence.  
 XX  
 KW ATM gene; ataxia-telangiectasia; progressive genetic disorder; antibody;  
 KW central nervous system; immune system; chromosomal instability; therapy;  
 KW cancer predisposition; radiation sensitivity; cell cycle abnormality;  
 KW multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;  
 KW general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;  
 KW blood vessel; bulbar conjunctiva; facial skin; A-T; mutein; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..8970  
 FT /tag= a  
 FT /transl\_except= (pos: 787..789, aa: Val)  
 FT misc\_difference 1406..1407  
 FT /tag= b  
 FT /note= "site of 201 nucleotide deletion"  
 XX  
 PN WO9636691-A1.  
 XX  
 XX 21-NOV-1996.  
 XX  
 XX 16-MAY-1996; 96WO-US07025.  
 XX  
 XX 08-APR-1996; 96US-0629001.  
 XX  
 XX 16-MAY-1995; 95US-0441822.  
 XX  
 XX (KOHN/) KOHN K I.  
 XX (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.  
 XX  
 PI Shiloh Y;  
 XX  
 DR WPI; 1997-012070/01.  
 DR P-PSDB; AAW19688.  
 XX  
 PT New isolated ataxia-telangiectasia gene - used to develop prods. for  
 PT the study, diagnosis and treatment of ataxia-telangiectasia.  
 XX  
 PS Claim 5; Page -; 153pp; English.  
 XX  
 CC AAT68731-T68785 represent mutations of the ATM gene of the invention  
 CC (see AAT43497 for wild type sequence). The ATM gene of the invention, is  
 CC located in the human genome at chromosome segment 11q22-23.  
 CC Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or  
 CC deletions (such as represented by this sequence) in the coding region of

CC the ATM gene. A-T is a progressive genetic disorder affecting the central  
 CC nervous and immune systems. A-T involves chromosomal instability, cancer  
 CC predisposition, radiation sensitivity, and cell cycle abnormalities. A-T  
 CC is a multi-system disease inherited in an autosomal recessive manner.  
 CC The wild type gene, can be used in methods for detecting carriers of a  
 CC defective gene that causes A-T. The gene can also be used to generate  
 CC antibodies. The methods and antibodies can be used in the study,  
 CC diagnosis and therapy of A-T.  
 XX

SQ Sequence 8970 BP; 2922 A; 1605 C; 1876 G; 2567 T; 0 other;

Query Match 64.8%; Score 16.2; DB 18; Length 8970;  
 Best Local Similarity 85.7%; Pred. No. 2.1e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 tcgcaagtcacagaattattt 25  
 || ||||| ||||| |||||  
 Db 1200 tcagaagtcacagaattattt 1220

RESULT 41  
 AAT68783  
 ID AAT68783 standard; DNA; 8996 BP.  
 XX  
 AC AAT68783;  
 XX

DT 20-AUG-1997 (first entry)

DE ATM mutant 4443del175 coding sequence.

XX  
 KW ATM gene; ataxia-telangiectasia; progressive genetic disorder; antibody;  
 KW central nervous system; immune system; chromosomal instability; therapy;  
 KW cancer predisposition; radiation sensitivity; cell cycle abnormality;  
 KW multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;  
 KW general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;  
 KW blood vessel; bulbar conjunctiva; facial skin; A-T; mutein; ss.  
 XX

OS Homo sapiens.

XX  
 FH Key Location/Qualifiers  
 FT CDS 1..4452  
 FT /tag= a  
 FT /transl\_except= (pos: 787..789, aa: Val)  
 FT /transl\_except= (pos: 1660..1662, aa: Ala)  
 FT /transl\_except= (pos: 1909..1911, aa: His)  
 FT /transl\_except= (pos: 2248..2250, aa: Asn)  
 FT misc\_difference 4442..4443  
 FT /tag= b  
 FT /note= "site of 175 nucleotide deletion"  
 XX

PN WO9636691-A1.

XX 21-NOV-1996.

XX 16-MAY-1996; 96WO-US07025.

XX 08-APR-1996; 96US-0629001.

XX 16-MAY-1995; 95US-0441822.

XX (KOHN/) KOHN K I.

XX (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.

XX Shiloh Y;

XX WPI; 1997-012070/01.

XX P-PSDB; AAW19700.

XX  
 PT New isolated ataxia-telangiectasia gene - used to develop prods. for  
 PT the study, diagnosis and treatment of ataxia-telangiectasia.

XX Claim 5; Page -; 153pp; English.

XX

CC AAT68731-T68785 represent mutations of the ATM gene of the invention  
 CC (see AAT43497 for wild type sequence). The ATM gene of the invention, is  
 CC located in the human genome at chromosome segment 11q22-23.  
 CC Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or  
 CC deletions (such as represented by this sequence) in the coding region of  
 CC the ATM gene. A-T is a progressive genetic disorder affecting the central  
 CC nervous and immune systems. A-T involves chromosomal instability, cancer  
 CC predisposition, radiation sensitivity, and cell cycle abnormalities. A-T  
 CC is a multi-system disease inherited in an autosomal recessive manner.  
 CC The wild type gene, can be used in methods for detecting carriers of a  
 CC defective gene that causes A-T. The gene can also be used to generate  
 CC antibodies. The methods and antibodies can be used in the study,  
 CC diagnosis and therapy of A-T.  
 XX  
 SQ Sequence 8996 BP; 2944 A; 1605 C; 1879 G; 2568 T; 0 other;

Query Match 64.8%; Score 16.2; DB 18; Length 8996;  
 Best Local Similarity 85.7%; Pred. No. 2.1e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 tcgcaagtcacagaattattt 25  
 || ||||| ||||| |||||  
 Db 1200 tcagaagtcacagaattattt 1220

RESULT 42  
 AAT68754  
 ID AAT68754 standard; DNA; 8996 BP.  
 XX  
 AC AAT68754;  
 DT 19-AUG-1997 (first entry)  
 XX  
 DE ATM mutant 4437del175 coding sequence.

XX ATM gene; ataxia-telangiectasia; progressive genetic disorder; antibody;  
 KW central nervous system; immune system; chromosomal instability; therapy;  
 KW cancer predisposition; radiation sensitivity; cell cycle abnormality;  
 KW multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;  
 KW general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;  
 KW blood vessel; bulbar conjunctiva; facial skin; A-T; mutein; ss.

OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..4452  
 FT /\*tag= a  
 FT /transl\_except= (pos: 787..789, aa: Val)  
 FT /transl\_except= (pos: 1660..1662, aa: Ala)  
 FT /transl\_except= (pos: 1909..1911, aa: His)  
 FT /transl\_except= (pos: 2248..2250, aa: Asn)  
 FT misc\_difference 4436..4437  
 FT /\*tag= b  
 FT /note= "site of 175 nucleotide deletion"

XX WO9636691-Al.  
 XX  
 PD 21-NOV-1996.  
 XX  
 PF 16-MAY-1996; 96WO-US07025.  
 XX  
 PR 08-APR-1996; 96US-0629001.  
 PR 16-MAY-1995; 95US-0441822.

XX (KOHN/) KOHN K I.  
 PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.  
 XX  
 PI Shiloh Y;  
 XX  
 DR WPI: 1997-012070/01.  
 DR P-PSDB; AAW19671.

XX

PT New isolated ataxia-telangiectasia gene - used to develop prods. for  
 PT the study, diagnosis and treatment of ataxia-telangiectasia.  
 XX  
 PS Claim 5; Page -: 153pp; English.  
 XX  
 CC AAT68731-T68785 represent mutations of the ATM gene of the invention  
 CC (see AAT43497 for wild type sequence). The ATM gene of the invention, is  
 CC located in the human genome at chromosome segment 11q22-23.  
 CC Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or  
 CC deletions (such as represented by this sequence) in the coding region of  
 CC the ATM gene. A-T is a progressive genetic disorder affecting the central  
 CC nervous and immune systems. A-T involves chromosomal instability, cancer  
 CC predisposition, radiation sensitivity, and cell cycle abnormalities. A-T  
 CC is a multi-system disease inherited in an autosomal recessive manner.  
 CC The wild type gene, can be used in methods for detecting carriers of a  
 CC defective gene that causes A-T. The gene can also be used to generate  
 CC antibodies. The methods and antibodies can be used in the study,  
 CC diagnosis and therapy of A-T.  
 XX  
 SQ Sequence 8996 BP; 2945 A; 1602 C; 1880 G; 2569 T; 0 other;

Query Match 64.8%; Score 16.2; DB 18; Length 8996;  
 Best Local Similarity 85.7%; Pred. No. 2.1e+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 tcgcaagtcacagaattattt 25  
 || ||||| ||||| |||||  
 Db 1200 tcagaagtcacagaattattt 1220

RESULT 43  
 AAT68774  
 ID AAT68774 standard; DNA; 8997 BP.  
 XX  
 AC AAT68774;  
 DT 20-AUG-1997 (first entry)  
 XX  
 DE ATM mutant 3403del174 coding sequence.

XX ATM gene; ataxia-telangiectasia; progressive genetic disorder; antibody;  
 KW central nervous system; immune system; chromosomal instability; therapy;  
 KW cancer predisposition; radiation sensitivity; cell cycle abnormality;  
 KW multi-system disease; autosomal recessive; cerebellar ataxia; cerebellum;  
 KW general motor dysfunction; Purkinje cell; oculocutaneous telangiectasia;  
 KW blood vessel; bulbar conjunctiva; facial skin; A-T; mutein; ss.

OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..8997  
 FT /\*tag= a  
 FT /transl\_except= (pos: 787..789, aa: Val)  
 FT /transl\_except= (pos: 1660..1662, aa: Ala)  
 FT /transl\_except= (pos: 1909..1911, aa: His)  
 FT /transl\_except= (pos: 2248..2250, aa: Asn)  
 FT misc\_difference 3402..3403  
 FT /\*tag= b  
 FT /note= "site of 174 nucleotide deletion"

XX WO9636691-Al.  
 XX  
 PD 21-NOV-1996.  
 XX  
 PF 16-MAY-1996; 96WO-US07025.  
 XX  
 PR 08-APR-1996; 96US-0629001.  
 PR 16-MAY-1995; 95US-0441822.

XX (KOHN/) KOHN K I.  
 PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.

XX





PD 21-NOV-1996.  
 XX  
 PF 16-MAY-1996; 96WO-US07025.  
 XX  
 PR 08-APR-1996; 96US-0629001.  
 PR 16-MAY-1995; 95US-0441822.  
 XX  
 PA (KOHN/) KOHN K I.  
 PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.  
 XX  
 PI Shiloh Y;  
 XX  
 DR WPI; 1997-012070/01.  
 DR P-PSDB; AAW19692.  
 XX  
 PT New isolated ataxia-telangiectasia gene - used to develop prods. for  
 PT the study, diagnosis and treatment of ataxia-telangiectasia.  
 XX  
 PS Claim 5; Page -: 153pp; English.  
 XX  
 CC AAT68731-T68785 represent mutations of the ATM gene of the invention  
 CC (see AAT43497 for wild type sequence). The ATM gene of the invention, is  
 CC located in the human genome at chromosome segment 11q22-23.  
 CC Ataxia-telangiectasia (A-T) is caused by mutations, insertions, or  
 CC deletions (such as represented by this sequence) in the coding region of  
 CC the ATM gene. A-T is a progressive genetic disorder affecting the central  
 CC nervous and immune systems. A-T involves chromosomal instability, cancer  
 CC predisposition, radiation sensitivity, and cell cycle abnormalities. A-T  
 CC is a multi-system disease inherited in an autosomal recessive manner.  
 CC The wild type gene, can be used in methods for detecting carriers of a  
 CC defective gene that causes A-T. The gene can also be used to generate  
 CC antibodies. The methods and antibodies can be used in the study,  
 CC diagnosis and therapy of A-T.  
 XX  
 SQ Sequence 9006 BP; 2932 A; 1610 C; 1891 G; 2573 T; 0 other;  
  
 Query Match 64.8%; Score 16.2; DB 18; Length 9006;  
 Best Local Similarity 85.7%; Pred. No. 2.le+02;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
 QY 5 tcgcaagtcacagaattattt 25  
 || |||||  
 Db 1200 tcagaagtcacagaatgattt 1220

Search completed: October 9, 2001, 11:36:16  
 Job time: 1691 sec

GenCore version 4.5  
Copyright (c). 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:39:28 ; Search time 221.13 Seconds  
(without alignments)  
21.403 Million cell updates/sec

Title: US-09-396-196F-2

Perfect score: 25

Sequence: 1 attgtcgcaagtcacagaattattt 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	17	68.0	17	2	US-07-944-259-2
6	16.8	67.2	3083	1	US-07-718-575-5
7	16.8	67.2	3083	1	US-08-481-206-5
8	16.8	67.2	3083	2	US-08-486-269A-5
9	16.2	64.8	500	4	US-08-642-274D-18
10	16.2	64.8	1374	1	US-08-123-761A-2
11	16.2	64.8	1771	1	US-09-166-203-48
12	16.2	64.8	5198	1	US-08-123-761A-1
13	16.2	64.8	9171	2	US-08-629-001A-2
14	16.2	64.8	9171	4	US-08-842-274D-2
15	16.2	64.8	9171	4	US-08-952-127-2
16	16.2	64.8	9385	2	US-08-874-266-1
17	16.2	64.8	9870	1	US-08-508-836A-9
18	16	64.0	3442	2	US-08-529-654-3
19	16	64.0	3459	2	US-08-980-060-3
20	16	64.0	3459	4	US-09-307-185-3
21	16	64.0	10968	2	US-08-680-327-2
22	16	64.0	10968	4	US-09-228-246-1
23	15.8	63.2	828	2	US-08-786-606-2
24	15.6	62.4	450	1	US-08-374-843B-4
25	15.6	62.4	450	1	US-08-374-843B-18
26	15.6	62.4	450	2	US-08-905-420-4
27	15.6	62.4	450	2	US-08-905-420-18

```

28 15.6 62.4 992 1 US-08-374-843B-9 Sequence 9, Appli
29 15.6 62.4 992 2 US-08-905-420-9 Sequence 9, Appli
30 15.6 62.4 997 1 US-08-374-843B-16 Sequence 16, Appli
31 15.6 62.4 997 2 US-08-905-420-16 Sequence 16, Appli
32 15.6 62.4 3348 1 US-08-222-616-35 Sequence 35, Appli
33 15.6 62.4 3348 5 PCT-US95-04228-35 Sequence 35, Appli
34 15.4 61.6 946 3 US-08-916-443A-15 Sequence 15, Appli
35 15.4 61.6 1070 3 US-08-904-284-1 Sequence 1, Appli
36 15.4 61.6 1116 3 US-08-916-443A-16 Sequence 16, Appli
37 15.4 61.6 2215 2 US-08-980-329C-1 Sequence 1, Appli
38 15.4 61.6 2437 3 US-08-804-284-6 Sequence 6, Appli
39 15.4 61.6 2946 1 US-08-687-379-1 Sequence 1, Appli
40 15.4 61.6 2946 1 US-08-687-379-3 Sequence 3, Appli
41 15.4 61.6 3220 1 US-08-254-573-1 Sequence 1, Appli
42 15.4 61.6 4368 5 PCT-US95-04567-3 Sequence 3, Appli
43 15.4 61.6 5994 3 US-09-032-365A-11 Sequence 11, Appli
44 15.4 61.6 8010 5 PCT-US95-11859-2 Sequence 2, Appli
45 15.4 61.6 9620 4 US-08-952-127-11 Sequence 11, Appli

```

## ALIGNMENTS

```

RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis In Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

```

Query Match 100.0%; Score 25; DB 2; Length 1041;  
 Best Local Similarity 100.0%; Pred. No. 0.003;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25  
 |||  
 DB 24 ATTGTCGCAAGTCACAGAATTATT 48

## RESULT 2

US-08-846-338-7  
 ; Sequence 7, Application US/08846338  
 ; Patent No. 5869719

## GENERAL INFORMATION:

APPLICANT: Patton, David  
 TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content  
 NUMBER OF SEQUENCES: 20  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: No. 5869719artis Corporation  
 STREET: 520 White Plains Road, P.O. Box 2005  
 CITY: Tarrytown  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10591

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30B  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/846,338  
 FILING DATE:

## CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:  
 NAME: Meigs, J. Timothy  
 REGISTRATION NUMBER: 38,241  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-541-8587  
 TELEFAX: 919-541-8689

## INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 1041 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1038  
 IDENTIFICATION METHOD: experimental  
 OTHER INFORMATION: /product= "biotin synthase"  
 OTHER INFORMATION: /evidence= EXPERIMENTAL

US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;  
 Best Local Similarity 100.0%; Pred. No. 0.003;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25  
 |||  
 DB 24 ATTGTCGCAAGTCACAGAATTATT 48

## RESULT 3

US-08-411-768B-1  
 ; Sequence 1, Application US/08411768B  
 ; Patent No. 6083712

## GENERAL INFORMATION:

APPLICANT: Olwen Birch  
 APPLICANT: Johann Brass  
 APPLICANT: Martin Fuhrmann

APPLICANT: Nicholas Shaw  
 TITLE OF INVENTION: Biotechnological Method  
 TITLE OF INVENTION: of Producing Biotin  
 NUMBER OF SEQUENCES: 19  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
 STREET: 30 Rockefeller Plaza  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10112

COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: WordPerfect  
 SOFTWARE: Version 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/411,768B  
 FILING DATE: 31-March-95  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: CH 3124/92  
 FILING DATE: 02-OCT-1992

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: CH 2134/93  
 FILING DATE: 15-JUL-1993

INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5872 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Escherichia coli  
 STRAIN: DSM498  
 IMMEDIATE SOURCE:  
 CLONE: pBO30A-15/9

FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 117..1157  
 IDENTIFICATION METHOD: experimental  
 OTHER INFORMATION: /codon\_start= 117  
 OTHER INFORMATION: /product= "Biotin synthase"  
 OTHER INFORMATION: /evidence= EXPERIMENTAL  
 OTHER INFORMATION: /gene= "bioB"  
 OTHER INFORMATION: /number= 1

FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 2295..3050  
 OTHER INFORMATION: /codon\_start= 2295  
 OTHER INFORMATION: /function= "involved in pimeloyl-CoA synthesis"  
 OTHER INFORMATION: /product= "protein"  
 OTHER INFORMATION: /gene= "bioC"  
 OTHER INFORMATION: /number= 3

FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 3750..5039  
 IDENTIFICATION METHOD: experimental  
 OTHER INFORMATION: /codon\_start= 3750  
 OTHER INFORMATION: /EC\_number= 2.6.1.62  
 OTHER INFORMATION: /product= "DAPA synthase"  
 OTHER INFORMATION: /evidence= EXPERIMENTAL  
 OTHER INFORMATION: /gene= "bioA"  
 OTHER INFORMATION: /number= 5  
 OTHER INFORMATION: /standard\_name= "S-Adenosyl-L-methionine:8-amino-7-oxononanoate aminotransf."

FEATURE:  
 NAME/KEY: CDS

```

; LOCATION: 5098..5574
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 5098
; OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
; OTHER INFORMATION: /product= "protein"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ORF1"
; OTHER INFORMATION: /number= 6
; FEATURE:
; NAME/KEY: -10_signal
; LOCATION: 45..49
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "promoter ptac"
; FEATURE:
; NAME/KEY: -35_signal
; LOCATION: 23..28
; OTHER INFORMATION: /standard_name= "promoter ptac"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 105..119
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "bioB RBS no.9"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 2284..2297
; OTHER INFORMATION: /standard_name= "bioC RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 3742..3752
; OTHER INFORMATION: /standard_name= "bioA RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 5088..5100
; OTHER INFORMATION: /standard_name= "ORF1 RBS"
; FEATURE:
; NAME/KEY: terminator
; LOCATION: 5583..5644
; OTHER INFORMATION: /standard_name= "rho-independent
; OTHER INFORMATION: transcriptional terminator"
; FEATURE:
; NAME/KEY: stem_loop
; LOCATION: 5583..5605
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 1..96
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /function= "promoter ptac"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: WO 87/01391 B1
; FILING DATE: 26-AUG-1986
; PUBLICATION DATE: 07-APR-1993
; US-08-411-768B-1

; Query Match 100.0%; Score 25; DB 3; Length 5872;
; Best Local Similarity 100.0%; Pred. No. 0.0038;
; Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 attgtcgaagtcacagaattattt 25
   |||||||
Db 140 ATTGCGCAAGTCACAGAATTATT 164

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass

```

```

; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: floppy
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411,768B
; FILING DATE: 31-March-95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 3124/92
; FILING DATE: 02-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 2134/93
; FILING DATE: 15-JUL-1993
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5872 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: DSM498
; IMMEDIATE SOURCE:
; CLONE: pBO30A15-9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1154..2308
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 1154
; OTHER INFORMATION: /EC_number= 2.3.1.47
; OTHER INFORMATION: /product= "KAPA synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "biof"
; OTHER INFORMATION: /number= 2
; OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3043..3753
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 3043
; OTHER INFORMATION: /EC_number= 6.3.3.3
; OTHER INFORMATION: /product= "PTB synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "bioD"
; OTHER INFORMATION: /number= 4
; OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 1141..1156
; OTHER INFORMATION: /standard_name= "bioF RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 3030..3045
; OTHER INFORMATION: /standard_name= "bioD RBS"
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: WO 87/01391 B1

```

; FILING DATE: 26-AUG-1986  
 ; PUBLICATION DATE: 07-APR-1993  
 US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;  
Best Local Similarity 100.0%; Pred. No. 0.0038;  
Matches 25: Conservative 0; Mismatches 0; Indels

Qy 1 attgtcgcaagtcacagaattattt 25  
140 ATTGTCGCAAGTCACAGAATTATTT 164

RESULT 5  
US-07-944-259-2/c  
; Sequence 2, Application US/07944259  
; Patent No. 5885792  
; GENERAL INFORMATION:  
; APPLICANT: Ifuku, Ohji  
; APPLICANT: Haze, Shinitiro  
; APPLICANT: Kishimoto, Jiro  
; APPLICANT: Nakahama, Kazuo  
; TITLE OF INVENTION: BIOTIN OPERON  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wegner, Cantor, Mueller & Player  
; STREET: 1233 20th Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036-8218  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/944,259  
; FILING DATE: 19920914  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cantor, Herbert I.  
; REGISTRATION NUMBER: 24,392  
; REFERENCE/DOCKET NUMBER: P-450-23557  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-887-0400  
; TELEFAX: 202-835-0605

Query Match	68.0%;	Score 17;	DB 2;	Length 17;
Best Local Similarity	100.0%;	Pred. No. 10;		
Matches	17.	Conservative	0:	Mismatches
			0:	Indels

yy 4 gtcgaagtcacagaat 20  
|||||  
bb 17 GTCGCAAGTCACAGAAT 1

RESULT 6  
US-07-718-575-5/c  
Sequence 5, Application US/07718575

```

1 Patent No. 5202257
2
3 GENERAL INFORMATION:
4
5 APPLICANT: Heinemann Ph.D., Stephen F.
6 APPLICANT: Boulter Ph.D., James R.
7 APPLICANT: Hollmann Ph.D., Michael NMN
8 APPLICANT: Bettler Ph.D., Bernhard NMN
9 APPLICANT: Jensen Ph.D., Jan E.
10
11 TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
12 METHODS
13
14 NUMBER OF SEQUENCES: 14
15
16 CORRESPONDENCE ADDRESS:
17
18 ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
19 STREET: 444 So. Flower St., Suite 2000
20 CITY: Los Angeles
21 STATE: California
22 COUNTRY: United States
23 ZIP: 90071-2921
24
25 COMPUTER READABLE FORM:
26
27 MEDIUM TYPE: Floppy disk
28
29 OPERATING SYSTEM: PC-DOS/MS-DOS
30
31 SOFTWARE: PatentIn Release #1.0, Version #1.25
32
33 CURRENT APPLICATION DATA:
34
35 APPLICATION NUMBER: US/07/718,575
36 FILING DATE: 19910813
37
38 CLASSIFICATION: 435
39
40 ATTORNEY/AGENT INFORMATION:
41
42 NAME: Reiter Ph.D., Stephen E.
43
44 REGISTRATION NUMBER: 31192
45
46 REFERENCE/DOCKET NUMBER: P31 8962
47
48 TELECOMMUNICATION INFORMATION:
49
50 TELEPHONE: (619) 535-9001
51 TELEFAX: (619) 535-8949
52
53 TELEX: 9103330318
54
55 INFORMATION FOR SEQ ID NO: 5:
56
57 SEQUENCE CHARACTERISTICS:
58
59 LENGTH: 3083 base pairs
60 TYPE: NUCLEIC ACID
61 STRANDEDNESS: single
62 TOPOLOGY: linear
63
64 MOLECULE TYPE: cDNA
65
66 IMMEDIATE SOURCE:
67
68 CLONE: GluR3
69
70 FEATURE:
71
72 NAME/KEY: CDS
73 LOCATION: 167..2833
74
75 IS-07-718-575-5

```

Query Match	67.2%	Score 16.8;	DB 1;	Length 3083;
Best Local Similarity	90.0%	Pred. No. 26;		
Matches 18: Conservative	0:	Mismatches	2:	Indels 0:
Gaps	0:			

Qy 1 attgtcgcaagtcacagaat 20  
|||||  
Db 3033 ATTCTCCCAAGTCTCAGACT 3014

RESULT 7  
US-08-481-206-5/c  
; Sequence 5, Application US/08481206  
; Patent No. 5739291  
; GENERAL INFORMATION:  
; APPLICANT: Heinemann Ph.D., Stephen F.  
; APPLICANT: Boulter Ph.D., James K.  
; APPLICANT: Hollmann Ph.D., Michael MNM  
; APPLICANT: Bettler Ph.D., Bernhard MNM  
; APPLICANT: Jensen Ph.D., Jan E.  
; TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretvy, Schroeder, Brueggemann & Clark

STREET: 444 So. Flower St., Suite 2000  
CITY: Los Angeles  
STATE: California  
COUNTRY: United States  
ZIP: 90071-2921  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/481,206  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/013,767  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Reiter Ph.D., Stephen E.  
REGISTRATION NUMBER: 31192  
REFERENCE/DOCKET NUMBER: P31 8962  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
TELEX: 9103330318  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3083 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
IMMEDIATE SOURCE:  
CLONE: Glur3  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 167...2833  
US-08-481-206-5

Query Match 67.2%; Score 16.8; DB 1; Length 3083;  
Best Local Similarity 90.0%; Pred. No. 26;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtcacagaat 20  
|||||  
Db 3033 ATGTGCGCAAGTCTCAGAGT 3014

RESULT 8  
US-08-486-269A-5/G  
Sequence 5, Application US/08486269A  
Patent No. 5945509  
GENERAL INFORMATION:  
APPLICANT: Heinemann, Stephen F.  
APPLICANT: Boulter, James R.  
APPLICANT: Hollmann, Michael  
APPLICANT: Bettler, Bernhard  
APPLICANT: Jensen, Jan E.  
TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Gray Cary Ware & Freidenrich LLP  
STREET: 4365 Executive Drive, Suite 1600  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows DEMONSTRATION Version 2.0D  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,269A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/013,767  
FILING DATE: 04-FEB-1993  
APPLICATION NUMBER: 07/718,575  
FILING DATE: 21-JUN-1991  
APPLICATION NUMBER: PCT/US90/06153  
FILING DATE: 25-OCT-1990  
APPLICATION NUMBER: 07/428,116  
FILING DATE: 27-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Reiter, Stephen E.  
REGISTRATION NUMBER: 31,192  
REFERENCE/DOCKET NUMBER: P41 9986  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-677-1409  
TELEFAX: 619-677-1465  
TELEX:  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3083 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
IMMEDIATE SOURCE:  
CLONE: Glur3  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 167...2830  
OTHER INFORMATION:  
US-08-486-269A-5

Query Match 67.2%; Score 16.8; DB 2; Length 3083;  
Best Local Similarity 90.0%; Pred. No. 26;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtcacagaat 20  
|||||  
Db 3033 ATGTGCGCAAGTCTCAGAGT 3014

RESULT 9  
US-08-642-274D-18  
Sequence 18, Application US/08642274D  
Patent No. 6200749  
GENERAL INFORMATION:  
APPLICANT: Shiloh, Yosef  
TITLE OF INVENTION: MUTATED FORMS OF THE ATAXIA-TELANGIECTASIA GENE AND METHOD TO  
FILE REFERENCE: 229000033  
CURRENT APPLICATION NUMBER: US/08/642,274D  
CURRENT FILING DATE: 1996-05-03  
NUMBER OF SEQ ID NOS: 220  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 18  
LENGTH: 500  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: genomic  
US-08-642-274D-18

Query Match 64.8%; Score 16.2; DB 4; Length 500;  
Best Local Similarity 85.7%; Pred. No. 39;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 tcgaagtcacagaattattt 25  
 || ||||| ||||| |||||  
 Db 245 tcgaagtcacagaattattt 265

## RESULT 10

US-08-123-761A-2/c  
 ; Sequence 2, Application US/08123761A  
 ; Patent No. 5589611  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Briggs, Steven P.  
 ; APPLICANT: Johal, Gurmukh S.  
 ; TITLE OF INVENTION: A DISEASE RESISTANCE GENE FROM MAIZE AS  
 ; TITLE OF INVENTION: A SELECTABLE MARKER  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.  
 ; STREET: 700 Capital Square, 400 Locust Street  
 ; CITY: Des Moines  
 ; STATE: Iowa  
 ; COUNTRY: USA  
 ; ZIP: 50309  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/123,761A  
 ; FILING DATE: 17-SEP-1993  
 ; CLASSIFICATION: 800  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Roth, Michael J.  
 ; REGISTRATION NUMBER: 29,342  
 ; REFERENCE/DOCKET NUMBER: 212-US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (515) 248-4844  
 ; TELEFAX: (515) 248-4844  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1374 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; US-08-123-761A-2

Query Match 64.8%; Score 16.2; DB 1; Length 1374;  
 Best Local Similarity 85.7%; Pred. No. 45;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaatt 21  
 ||| ||||| ||||| |||||  
 Db 1327 ATTATCGCAAGTCATAAAATT 1307

## RESULT 11

US-09-166-203-48  
 ; Sequence 48, Application US/09166203A  
 ; Patent No. 5968826  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bennett, C. Frank  
 ; APPLICANT: Condon, Tom P.  
 ; TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION  
 ; FILE REFERENCE: ISPH-0323  
 ; CURRENT APPLICATION NUMBER: US/09/166,203A  
 ; CURRENT FILING DATE: 1998-10-05  
 ; NUMBER OF SEQ ID NOS: 60  
 ; SEQ ID NO 48  
 ; LENGTH: 1771  
 ; TYPE: DNA

; ORGANISM: Mus musculus  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1193)..(1387)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1709)..(1771)  
 ; PUBLICATION INFORMATION:  
 ; DATABASE ACCESSION NUMBER: L20788 Genbank  
 ; DATABASE ENTRY DATE: 1996-04-18  
 ; US-09-166-203-48

Query Match 64.8%; Score 16.2; DB 2; Length 1771;  
 Best Local Similarity 85.7%; Pred. No. 47;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaatt 21  
 ||| ||||| ||||| |||||  
 Db 306 agtgaggcaagtcacagaatt 326

## RESULT 12

US-08-123-761A-1/c  
 ; Sequence 1, Application US/08123761A  
 ; Patent No. 5589611  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Briggs, Steven P.  
 ; APPLICANT: Johal, Gurmukh S.  
 ; TITLE OF INVENTION: A DISEASE RESISTANCE GENE FROM MAIZE AS  
 ; TITLE OF INVENTION: A SELECTABLE MARKER  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.  
 ; STREET: 700 Capital Square, 400 Locust Street  
 ; CITY: Des Moines  
 ; STATE: Iowa  
 ; COUNTRY: USA  
 ; ZIP: 50309  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/123,761A  
 ; FILING DATE: 17-SEP-1993  
 ; CLASSIFICATION: 800  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Roth, Michael J.  
 ; REGISTRATION NUMBER: 29,342  
 ; REFERENCE/DOCKET NUMBER: 212-US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (515) 248-4800  
 ; TELEFAX: (515) 248-4844  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 5198 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-08-123-761A-1

Query Match 64.8%; Score 16.2; DB 1; Length 5198;  
 Best Local Similarity 85.7%; Pred. No. 54;  
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaatt 21  
 ||| ||||| ||||| |||||  
 Db 4889 ATTATCGCAAGTCATAAAATT 4869



```
RESULT 13
US-08-629-001A-2
; Sequence 2, Application US/08629001A
; Patent No. 5858661
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; TITLE OF INVENTION: ATAXIA-TELANGELECTASIA GENE AND ITS
; TITLE OF INVENTION: GENOMIC ORGANIZATION
; NUMBER OF SEQUENCES: 139
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kohn & Associates
; STREET: 30500 No. 5858661thwestern Hwy.
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: US
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/629,001A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: 2290.00032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 539-5050
; TELEFAX: (810) 539-5055
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 11q22-23
;
US-08-629-001A-2

Query Match 64.8%; Score 16.2; DB 2; Length 9171;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 tcgcaagtcacagaattattt 25
|| ||||| ||||| |||||
DB 1200 TCAGAAGTCACAGAATGATT 1220

RESULT 14
US-08-642-274D-2
; Sequence 2, Application US/08642274D
; Patent No. 6200749
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; TITLE OF INVENTION: MUTATED FORMS OF THE ATAXIA-TELANGELECTASIA GENE AND METHOD TO
; TITLE OF INVENTION: SCREEN FOR A PARTIAL A-T PHENOTYPE
; FILE REFERENCE: 229000033
; CURRENT APPLICATION NUMBER: US/08/642,274D
; CURRENT FILING DATE: 1996-05-03
; NUMBER OF SEQ ID NOS: 220
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 9171
; TYPE: DNA
; ORGANISM: Homo Sapiens
```

```
; FEATURE:
; OTHER INFORMATION:
US-08-642-274D-2

Query Match 64.8%; Score 16.2; DB 4; Length 9171;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 tcgcaagtcacagaattattt 25
|| ||||| ||||| |||||
DB 1200 tcagaagtcacagaatgattt 1220

RESULT 15
US-08-952-127-2
; Sequence 2, Application US/08952127
; Patent No. 6211336
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; APPLICANT: Tagle, Danilo A.
; APPLICANT: Collins, Francis S.
; TITLE OF INVENTION: ATAXIA-TELANGELECTASIA GENE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kohn & Associates
; STREET: 30500 No. 6211336thwestern Hwy., Suite 410
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: U.S.
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/952,127
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,995
; REFERENCE/DOCKET NUMBER: 2290.00029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 810-539-5050
; TELEFAX: 810-539-5055
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 11q22-23
;
US-08-952-127-2

Query Match 64.8%; Score 16.2; DB 4; Length 9171;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 tcgcaagtcacagaattattt 25
|| ||||| ||||| |||||
DB 1200 TCAGAAGTCACAGAATGATT 1220

RESULT 16
US-08-874-266-1
; Sequence 1, Application US/08874266
```

Patent No. 5955279  
GENERAL INFORMATION:  
APPLICANT: Gatti, Richard A.  
TITLE OF INVENTION: ATAXIA-TELANGIECTASIA: MUTATIONS IN THE ATM GENE  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson and Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/874,266  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ways Vensko, Nancy  
REGISTRATION NUMBER: 36,298  
REFERENCE/DOCKET NUMBER: UCLA006.006A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-235-8550  
TELEFAX: 619-235-0176  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9385 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 190...9357  
OTHER INFORMATION:  
US-08-874-266-1

Query Match 64.8%; Score 16.2; DB 2; Length 9385;  
Best Local Similarity 85.7%; Pred. No. 59;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 tcgcaagtcacagaattattt 25  
|| ||||| ||||| |||||  
Db 1389 TCAGAAGTCACAGAATGATT 1409

RESULT 17  
US-08-836A-9  
Sequence 9, Application US/08508836A  
Patent No. 5777093  
GENERAL INFORMATION:  
APPLICANT: Shiloh, Yosef  
APPLICANT: Taglie, Danilo A.  
APPLICANT: Collins, Francis S.  
TITLE OF INVENTION: Ataxia-Telangiectasia Gene  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Reising, Ethington, Barnard & Perry  
STREET: P.O. Box 4390  
CITY: Troy  
STATE: Michigan  
COUNTRY: US  
ZIP: 48099  
COMPUTER READABLE FORM:

Patent No. 5955279  
GENERAL INFORMATION:  
APPLICANT: Gatti, Richard A.  
TITLE OF INVENTION: ATAXIA-TELANGIECTASIA: MUTATIONS IN THE ATM GENE  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson and Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/874,266  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ways Vensko, Nancy  
REGISTRATION NUMBER: 36,298  
REFERENCE/DOCKET NUMBER: UCLA006.006A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-235-8550  
TELEFAX: 619-235-0176  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9385 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 190...9357  
OTHER INFORMATION:  
US-08-874-266-1

Query Match 64.8%; Score 16.2; DB 2; Length 9385;  
Best Local Similarity 85.7%; Pred. No. 59;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 tcgcaagtcacagaattattt 25  
|| ||||| ||||| |||||  
Db 1389 TCAGAAGTCACAGAATGATT 1409

RESULT 17  
US-08-836A-9  
Sequence 9, Application US/08508836A  
Patent No. 5777093  
GENERAL INFORMATION:  
APPLICANT: Shiloh, Yosef  
APPLICANT: Taglie, Danilo A.  
APPLICANT: Collins, Francis S.  
TITLE OF INVENTION: Ataxia-Telangiectasia Gene  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Reising, Ethington, Barnard & Perry  
STREET: P.O. Box 4390  
CITY: Troy  
STATE: Michigan  
COUNTRY: US  
ZIP: 48099  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/508,836A  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Kohn, Kenneth I.  
REGISTRATION NUMBER: 30,955  
REFERENCE/DOCKET NUMBER: P-313 (TAU)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (810) 689-3500  
TELEFAX: (810) 689-4071  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9870 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-508-836A-9

Query Match 64.8%; Score 16.2; DB 1; Length 9870;  
Best Local Similarity 85.7%; Pred. No. 59;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 tcgcaagtcacagaattattt 25  
|| ||||| ||||| |||||  
Db 1364 TCAGAAGTCACAGAATGATT 1384

RESULT 18  
US-08-529-654-3/c  
Sequence 3, Application US/08529654  
Patent No. 5739284  
GENERAL INFORMATION:  
APPLICANT: HEDIGER, MATTHIAS  
APPLICANT: KANAL, YOSHIKATSU  
TITLE OF INVENTION: COMPOSITIONS CORRESPONDING TO A HIGH  
AFFINITY GLUTAMATE TRANSPORTER MOLECULE AND METHODS FOR  
TITLE OF INVENTION: AFFINITY GLUTAMATE TRANSPORTER MOLECULE AND METHODS FOR  
TITLE OF INVENTION: MAKING AND USING SAME  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
STREET: 600 ATLANTIC AVENUE  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/529,654  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/194,719  
FILING DATE: 10-FEB-1994  
APPLICATION NUMBER: US 07/965,676  
FILING DATE: 19-OCT-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: JANIUK, ANTHONY J.  
REGISTRATION NUMBER: 29,809  
REFERENCE/DOCKET NUMBER: B0801/7021  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441  
INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3442 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA to mRNA  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: LEPORIDAE (RABBIT)  
 US-08-529-654-3

Query Match 64.0%; Score 16; DB 1; Length 3442;  
 Best Local Similarity 79.2%; Pred. No. 64;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattatt 25  
 Db 2117 TTGTAACCAATGCACAGAATTATT 2094

## RESULT 19

US-08-980-060-3  
 ; Sequence 3, Application US/08980060  
 ; Patent No. 5965421

; GENERAL INFORMATION:  
 ; APPLICANT: NI, JIAN  
 ; APPLICANT: FENG, PING  
 ; APPLICANT: MUZIO, MARTA  
 ; APPLICANT: DIXIT, VISHVA M.  
 ; TITLE OF INVENTION: HUMAN IRAK-2  
 ; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
 ; STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600  
 ; CITY: WASHINGTON  
 ; STATE: D.C.

; COUNTRY: USA  
 ; ZIP: 20005-3934

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/980,060

; FILING DATE: Herewith  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: STEFFE, ERIC K.

; REGISTRATION NUMBER: 36,688  
 ; REFERENCE/DOCKET NUMBER: 1488.1010000/EKS/AJK  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3459 base pairs

; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS

; LOCATION: 34..1908  
 US-08-980-060-3

Query Match 64.0%; Score 16; DB 2; Length 3459;  
 Best Local Similarity 79.2%; Pred. No. 64;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattatt 24  
 Db 2607 AATCTTGAATTCACATAATTATT 2630

## RESULT 20

US-09-307-185-3  
 ; Sequence 3, Application US/09307185  
 ; Patent No. 6222019

; GENERAL INFORMATION:  
 ; APPLICANT: NI, JIAN  
 ; APPLICANT: FENG, PING  
 ; APPLICANT: MUZIO, MARTA  
 ; APPLICANT: DIXIT, VISHVA M.  
 ; TITLE OF INVENTION: HUMAN IRAK-2  
 ; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
 ; STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600  
 ; CITY: WASHINGTON  
 ; STATE: D.C.

; COUNTRY: USA  
 ; ZIP: 20005-3934

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/307,185

; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/980,060

; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: STEFFE, ERIC K.

; REGISTRATION NUMBER: 36,688  
 ; REFERENCE/DOCKET NUMBER: 1488.1010000/EKS/AJK  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3459 base pairs

; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS

; LOCATION: 34..1908  
 US-09-307-185-3

Query Match 64.0%; Score 16; DB 4; Length 3459;  
 Best Local Similarity 79.2%; Pred. No. 64;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattatt 24  
 Db 2607 AATCTTGAATTCACATAATTATT 2630

## RESULT 21

US-08-680-327-2/c  
 ; Sequence 2, Application US/08680327  
 ; Patent No. 5859321

; GENERAL INFORMATION:  
 ; APPLICANT: Staskawicz, Brian S., Oldroyd, Giles Edward,  
 ; APPLICANT: Salmeron, John M., Rommens, Caius  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PLANT  
 ; TITLE OF INVENTION: PATHOGEN RESISTANCE

NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klarquist Sparkman Campbell Leigh &  
ADDRESSEE: Winston  
STREET: One World Trade Center  
STREET: 121 S.W. Salmon Street  
STREET: Suite 1600  
CITY: Portland  
STATE: Oregon  
COUNTRY: United States of America  
ZIP: 97204  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Disk, 3-1/2 inch  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: MS DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/680,327  
FILING DATE: July 11, 1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/310,912  
FILING DATE: September 22, 1994  
CLASSIFICATION: 800  
APPLICATION NUMBER: 08/227,360  
FILING DATE: April 13, 1994  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Dow, Alan. E.  
REGISTRATION NUMBER: 35,123  
REFERENCE/DOCKET NUMBER: 5151-45038  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (503) 226-7391  
TELEFAX: (503) 228-9446  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10968 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double stranded  
TOPOLOGY: linear  
US-08-680-327-2

Query Match 64.0%; Score 16; DB 2; Length 10968;  
Best Local Similarity 79.2%; Pred. No. 75;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattatt 25  
||||| ||||| | |||||  
Db 745 TTGTGCAAGTCAAAATTGTTATT 722

RESULT 22  
US-09-228-246-1/c  
Sequence 1, Application US/09228246  
Patent No. 6245510  
GENERAL INFORMATION:  
APPLICANT: Staskawicz, B. S. et al.  
TITLE OF INVENTION: PRP Protein and Nucleic Acid Sequences: Compositions  
FILE REFERENCE: 51700  
CURRENT APPLICATION NUMBER: US/09/228,246  
CURRENT FILING DATE: 1999-01-11  
EARLIER APPLICATION NUMBER: 08/680,327  
EARLIER FILING DATE: 1996-07-11  
EARLIER APPLICATION NUMBER: 08/310,912  
EARLIER FILING DATE: 1994-09-22  
EARLIER APPLICATION NUMBER: 08/227,360  
EARLIER FILING DATE: 1994-04-13  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 1  
LENGTH: 10968

TYPE: DNA  
ORGANISM: Lycopersicon esculentum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (3879)..(8186)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (8300)..(9466)  
US-09-228-246-1

Query Match 64.0%; Score 16; DB 4; Length 10968;  
Best Local Similarity 79.2%; Pred. No. 75;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattatt 25  
||||| ||||| | |||||  
Db 745 TTGTGCAAGTCAAAATTGTTATT 722

RESULT 23  
US-08-786-606-2  
Sequence 2, Application US/08786606  
Patent No. 5861495  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Au-Young, Janice  
APPLICANT: Coleman, Roger  
APPLICANT: Goli, Surya K.  
TITLE OF INVENTION: NOVEL HUMAN ZINC-BINDING  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/786,606  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy RJ  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0173 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 828 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-786-606-2

Query Match 63.2%; Score 15.8; DB 2; Length 828;  
Best Local Similarity 89.5%; Pred. No. 65;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 gcaagtcacagaattatt 25

```

Db 90 GAAAGTCACAGTAATT 108
      | ||||| ||||| |||||
RESULT 24
US-08-374-843B-4
; Sequence 4, Application US/08374843B
; Patent No. 5726016
; GENERAL INFORMATION:
; APPLICANT: Demuth, Donald R.
; APPLICANT: Lally, Edward T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
; OF DISEASES ASSOCIATED WITH ACTINOBACILLUS
; TITLE OF INVENTION: OF DISEASES ASSOCIATED WITH ACTINOBACILLUS
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Street
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,843B
; FILING DATE: 18-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary, Kathryn Ph.D.
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 567-2020
; TELEFAX: (215) 567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 450 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-374-843B-4

Query Match 62.4%; Score 15.6; DB 1; Length 450;
Best Local Similarity 81.8%; Pred. No. 74;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattat 23
      ||||| ||||| |||||
Db 260 TTGTCGCAAGTCCATAGTTAT 281

RESULT 25
US-08-374-843B-18
; Sequence 18, Application US/08374843B
; Patent No. 5726016
; GENERAL INFORMATION:
; APPLICANT: Demuth, Donald R.
; APPLICANT: Lally, Edward T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
; OF DISEASES ASSOCIATED WITH ACTINOBACILLUS
; TITLE OF INVENTION: ACTINOMYCETEMCOMITANS INFECTION
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Street
; CITY: Philadelphia

```

```

; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,843B
; FILING DATE: 18-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary, Kathryn Ph.D.
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 567-2020
; TELEFAX: (215) 567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 450 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-374-843B-18

Query Match 62.4%; Score 15.6; DB 1; Length 450;
Best Local Similarity 81.8%; Pred. No. 74;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattat 23
      ||||| ||||| |||||
Db 260 TTGTCGCAAGTCCATAGTTAT 281

RESULT 26
US-08-905-420-4
; Sequence 4, Application US/08905420
; Patent No. 5861255
; GENERAL INFORMATION:
; APPLICANT: Demuth, Donald R.
; APPLICANT: Lally, Edward T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
; OF DISEASES ASSOCIATED WITH ACTINOBACILLUS
; TITLE OF INVENTION: ACTINOMYCETEMCOMITANS INFECTION
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,420
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: US 08/374,843
; FILING DATE: 18-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary, Kathryn Ph.D.
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)

```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 567-2020  
TELEFAX: (215) 567-2991  
TELEX: 831-494  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 450 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-905-420-4

Query Match 62.4%; Score 15.6; DB 2; Length 450;  
Best Local Similarity 81.8%; Pred. No. 74;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 ttgtcgcaagtcacagaattat 23  
||||| ||| |||  
Db 260 TTGTCGCAAGTGCCATAGTTAT 281

RESULT 27  
US-08-905-420-18  
; Sequence 18, Application US/08905420  
; Patent No. 5861255  
; GENERAL INFORMATION:  
; APPLICANT: Demuth, Donald R.  
; APPLICANT: Lally, Edward T.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS  
; TITLE OF INVENTION: OF DISEASES ASSOCIATED WITH ACTINOBACILLUS  
; TITLE OF INVENTION: ACTINOMYCETEMCOMITANS INFECTION  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.  
; STREET: 1601 Market Street, 36th Floor  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19103-2398  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/905,420  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/374,843  
; FILING DATE: 18-JAN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Leary, Kathryn Ph.D.  
; REGISTRATION NUMBER: 36,317  
; REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 567-2020  
; TELEFAX: (215) 567-2991  
; TELEX: 831-494  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 450 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-905-420-18

Query Match 62.4%; Score 15.6; DB 2; Length 450;  
Best Local Similarity 81.8%; Pred. No. 74;

Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 ttgtcgcaagtcacagaattat 23  
||||| ||| |||  
Db 260 TTGTCGCAAGTGCCATAGTTAT 281

RESULT 28  
US-08-374-843B-9  
; Sequence 9, Application US/08374843B  
; Patent No. 5726016  
; GENERAL INFORMATION:  
; APPLICANT: Demuth, Donald R.  
; APPLICANT: Lally, Edward T.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS  
; TITLE OF INVENTION: OF DISEASES ASSOCIATED WITH ACTINOBACILLUS  
; TITLE OF INVENTION: ACTINOMYCETEMCOMITANS INFECTION  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.  
; STREET: 1601 Market Street, 36th Street  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19103-2398  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/374,843B  
; FILING DATE: 18-JAN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Leary, Kathryn Ph.D.  
; REGISTRATION NUMBER: 36,317  
; REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 567-2020  
; TELEFAX: (215) 567-2991  
; TELEX: 831-494  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 992 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 200..652  
US-08-374-843B-9

Query Match 62.4%; Score 15.6; DB 1; Length 992;  
Best Local Similarity 81.8%; Pred. No. 82;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 ttgtcgcaagtcacagaattat 23  
||||| ||| |||  
Db 459 TTGTCGCAAGTGCCATAGTTAT 480

RESULT 29  
US-08-905-420-9  
; Sequence 9, Application US/08905420  
; Patent No. 5861255  
; GENERAL INFORMATION:  
; APPLICANT: Demuth, Donald R.  
; APPLICANT: Lally, Edward T.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS  
; TITLE OF INVENTION: OF DISEASES ASSOCIATED WITH ACTINOBACILLUS

;/ TITLE OF INVENTION: ACTINOMYCETEMCOMITANS INFECTION  
;/ NUMBER OF SEQUENCES: 29  
;/ CORRESPONDENCE ADDRESS:  
;/ ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.  
;/ STREET: 1601 Market Street, 36th Floor  
;/ CITY: Philadelphia  
;/ STATE: Pennsylvania  
;/ COUNTRY: USA  
;/ ZIP: 19103-2398  
;/ COMPUTER READABLE FORM:  
;/ MEDIUM TYPE: Floppy disk  
;/ COMPUTER: IBM PC compatible  
;/ OPERATING SYSTEM: PC-DOS/MS-DOS  
;/ SOFTWARE: PatentIn Release #1.0, Version #1.30  
;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: US/08/905,420  
;/ FILING DATE:  
;/ CLASSIFICATION: 435  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: US 08/374,843  
;/ FILING DATE: 18-JAN-1995  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: Leary, Kathryn Ph.D.  
;/ REGISTRATION NUMBER: 36,317  
;/ REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: (215) 567-2020  
;/ TELEFAX: (215) 567-2991  
;/ TELEX: 831-494  
;/ INFORMATION FOR SEQ ID NO: 9:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 992 base pairs  
;/ TYPE: nucleic acid  
;/ STRANDEDNESS: single  
;/ TOPOLOGY: linear  
;/ MOLECULE TYPE: DNA (genomic)  
;/ FEATURE:  
;/ NAME/KEY: CDS  
;/ LOCATION: 200..652  
;/ US-08-905-420-9

Query Match 62.4%; Score 15.6; DB 2; Length 992;  
Best Local Similarity 81.8%; Pred. No. 82;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttgtcgcaagtcacagaattat 23  
||||| ||| |||  
Db 459 TTGTGCGCAAGTCCCATAGTTAT 480

RESULT 30  
US-08-374-843B-16  
;/ Sequence 16, Application US/08374843B  
;/ Patent No. 5726016  
;/ GENERAL INFORMATION:  
;/ APPLICANT: Demuth, Donald R.  
;/ APPLICANT: Lally, Edward T.  
;/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS  
;/ TITLE OF INVENTION: OF DISEASES ASSOCIATED WITH ACTINOBACILLUS  
;/ TITLE OF INVENTION: ACTINOMYCETEMCOMITANS INFECTION  
;/ NUMBER OF SEQUENCES: 29  
;/ CORRESPONDENCE ADDRESS:  
;/ ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.  
;/ STREET: 1601 Market Street, 36th Street  
;/ CITY: Philadelphia  
;/ STATE: Pennsylvania  
;/ COUNTRY: USA  
;/ ZIP: 19103-2398  
;/ COMPUTER READABLE FORM:  
;/ MEDIUM TYPE: Floppy disk  
;/ COMPUTER: IBM PC compatible  
;/ OPERATING SYSTEM: PC-DOS/MS-DOS

;/ SOFTWARE: PatentIn Release #1.0, Version #1.30  
;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: US/08/374,843B  
;/ FILING DATE: 18-JAN-1995  
;/ CLASSIFICATION: 435  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: Leary, Kathryn Ph.D.  
;/ REGISTRATION NUMBER: 36,317  
;/ REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: (215) 567-2020  
;/ TELEFAX: (215) 567-2991  
;/ TELEX: 831-494  
;/ INFORMATION FOR SEQ ID NO: 16:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 997 base pairs  
;/ TYPE: nucleic acid  
;/ STRANDEDNESS: single  
;/ TOPOLOGY: linear  
;/ MOLECULE TYPE: DNA (genomic)  
;/ US-08-374-843B-16

Query Match 62.4%; Score 15.6; DB 1; Length 997;  
Best Local Similarity 81.8%; Pred. No. 82;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttgtcgcaagtcacagaattat 23  
||||| ||| |||  
Db 459 TTGTGCGCAAGTCCCATAGTTAT 480

RESULT 31  
US-08-905-420-16  
;/ Sequence 16, Application US/08905420  
;/ Patent No. 5861255  
;/ GENERAL INFORMATION:  
;/ APPLICANT: Demuth, Donald R.  
;/ APPLICANT: Lally, Edward T.  
;/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS  
;/ TITLE OF INVENTION: OF DISEASES ASSOCIATED WITH ACTINOBACILLUS  
;/ TITLE OF INVENTION: ACTINOMYCETEMCOMITANS INFECTION  
;/ NUMBER OF SEQUENCES: 29  
;/ CORRESPONDENCE ADDRESS:  
;/ ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.  
;/ STREET: 1601 Market Street, 36th Floor  
;/ CITY: Philadelphia  
;/ STATE: Pennsylvania  
;/ COUNTRY: USA  
;/ ZIP: 19103-2398  
;/ COMPUTER READABLE FORM:  
;/ MEDIUM TYPE: Floppy disk  
;/ COMPUTER: IBM PC compatible  
;/ OPERATING SYSTEM: PC-DOS/MS-DOS  
;/ SOFTWARE: PatentIn Release #1.0, Version #1.30  
;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: US/08/905,420  
;/ FILING DATE:  
;/ CLASSIFICATION: 435  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: US 08/374,843  
;/ FILING DATE: 18-JAN-1995  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: Leary, Kathryn Ph.D.  
;/ REGISTRATION NUMBER: 36,317  
;/ REFERENCE/DOCKET NUMBER: 9596-7 (F-1080)  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: (215) 567-2020  
;/ TELEFAX: (215) 567-2991  
;/ TELEX: 831-494  
;/ INFORMATION FOR SEQ ID NO: 16:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 997 base pairs

; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 US-08-905-420-16

Query Match 62.4%; Score 15.6; DB 2; Length 997;  
 Best Local Similarity 81.8%; Pred. No. 82;  
 Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattat 23  
 ||||| ||||| ||||| |||||  
 Db 459 TTGTCGCAAGTCCCATAGTTAT 480

## RESULT 32

US-08-222-616-35/c  
 ; Sequence 35, Application US/08222616  
 ; Patent No. 5635177

## GENERAL INFORMATION:

; APPLICANT: Bennett, Brian D.  
 ; APPLICANT: Goeddel, David  
 ; APPLICANT: Lee, James M.  
 ; APPLICANT: Matthews, William  
 ; APPLICANT: Tsai, Siao Ping  
 ; APPLICANT: Wood, William I.  
 ; TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST  
 ; TITLE OF INVENTION: ANTIBODIES  
 ; NUMBER OF SEQUENCES: 42  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 460 Point San Bruno Blvd  
 ; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080

## COMPUTER READABLE FORM:

; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: patin (Genentech)

; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/222.616  
 ; FILING DATE: 4-APR-1994

## CLASSIFICATION:

; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US93/00586  
 ; FILING DATE: 22-JAN-1993

## PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/826935  
 ; FILING DATE: 22-JAN-1992

## ATTORNEY/AGENT INFORMATION:

; NAME: Lee, Wendy M.  
 ; REGISTRATION NUMBER:  
 ; REFERENCE/DOCKET NUMBER: 821P2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415/225-1994

; TELEFAX: 415/952-9881

; TELEX: 910/371-7168

## INFORMATION FOR SEQ ID NO: 35:

## SEQUENCE CHARACTERISTICS:

; LENGTH: 3348 bases

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-222-616-35

Query Match 62.4%; Score 15.6; DB 1; Length 3348;  
 Best Local Similarity 81.8%; Pred. No. 98;  
 Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattat 23  
 ||||| ||||| ||||| |||||  
 Db 1974 TGGTCGAAAAGTCTCAGAACTAT 1953

## RESULT 33

PCT-US95-04228-35/c  
 ; Sequence 35, Application PC/TUS9504228  
 ; GENERAL INFORMATION:

; APPLICANT: Genentech, Inc.  
 ; APPLICANT: Bennett, Brian D.  
 ; APPLICANT: Goeddel, David  
 ; APPLICANT: Lee, James M.  
 ; APPLICANT: Matthews, William  
 ; APPLICANT: Tsai, Siao Ping  
 ; APPLICANT: Wood, William I.

## TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST ANTIBODIES

; NUMBER OF SEQUENCES: 45

## CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.  
 ; STREET: 460 Point San Bruno Blvd  
 ; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080

## COMPUTER READABLE FORM:

; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: patin (Genentech)

; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US95/04228

## FILING DATE:

; CLASSIFICATION:

## PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/222616

; FILING DATE: 04-APR-1994

## ATTORNEY/AGENT INFORMATION:

; NAME: Wendy M. Lee  
 ; REGISTRATION NUMBER: 00.000  
 ; REFERENCE/DOCKET NUMBER: 821P3PCT  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415/225-1994

; TELEFAX: 415/952-9881

; TELEX: 910/371-7168

## INFORMATION FOR SEQ ID NO: 35:

## SEQUENCE CHARACTERISTICS:

; LENGTH: 3348 bases

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

PCT-US95-04228-35

Query Match 62.4%; Score 15.6; DB 5; Length 3348;  
 Best Local Similarity 81.8%; Pred. No. 98;  
 Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattat 23  
 ||||| ||||| ||||| |||||  
 Db 1974 TGGTCGAAAAGTCTCAGAACTAT 1953

## RESULT 34

US-08-916-443A-15  
 ; Sequence 15, Application US/08916443A  
 ; Patent No. 6001986

## GENERAL INFORMATION:

; APPLICANT: Yong Sig KIM  
 ; APPLICANT: Sun Chung PARK  
 ; APPLICANT: Soo Kyung OH  
 ; APPLICANT: Hosull LEE  
 ; APPLICANT: Jeong Woo CHO







```
;
; GENERAL INFORMATION:
; APPLICANT: Hoeger, Thomas
; APPLICANT: Ultsch, Andreas
; APPLICANT: Bach, Alfred
; APPLICANT: Sterrer, Sylvia
; APPLICANT: Lemaire, Hans-Georg
; TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
; PREPARATION AND THEIR USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kell & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,379
; FILING DATE: 05-AUG-1996
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2946 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORGANISM: Homo sapiens
; DEVELOPMENTAL STAGE: Adult
; TISSUE TYPE: Brain
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 144..2861
; US-08-687-379-1

Query Match 61.6%; Score 15.4; DB 1; Length 2946;
Best Local Similarity 94.1%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacaga 18
||||||| |||||
Db 279 TTGTCGCAACTCACAGA 295

RESULT 40
US-08-687-379-3
; Sequence 3, Application US/08687379
; Patent No. 5756697
; GENERAL INFORMATION:
; APPLICANT: Hoeger, Thomas
; APPLICANT: Ultsch, Andreas
; APPLICANT: Bach, Alfred
; APPLICANT: Sterrer, Sylvia
; APPLICANT: Lemaire, Hans-Georg
; TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
; PREPARATION AND THEIR USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kell & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
```

```
;
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,379
; FILING DATE: 05-AUG-1996
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2946 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORGANISM: Homo sapiens
; DEVELOPMENTAL STAGE: Adult
; TISSUE TYPE: Brain
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 144..2861
; US-08-687-379-3

Query Match 61.6%; Score 15.4; DB 1; Length 2946;
Best Local Similarity 94.1%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacaga 18
||||||| |||||
Db 279 TTGTCGCAACTCACAGA 295

RESULT 41
US-08-254-573-1
; Sequence 1, Application US/08254573
; Patent No. 5610032
; GENERAL INFORMATION:
; APPLICANT: KAMBOJ, Rajender
; APPLICANT: ELIOTT, Candace
; APPLICANT: NUTT, Stephen
; TITLE OF INVENTION: AMPA-BINDING HUMAN GLUR1 RECEPTORS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/254,573
; FILING DATE: 06-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/896,611
; FILING DATE: 10-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16777/179 ALLE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3220 base pairs
```

```

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 62..2782
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 62..115
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 116..2782
;
US-08-254-573-1

Query Match 61.6%; Score 15.4; DB 1; Length 3220;
Best Local Similarity 94.1%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacaga 18
   ||||| |||||
Db 197 TTGTGCAACTCACAGA 213

RESULT 42
PCT-US95-04567-3
; Sequence 3, Application PC/TUS9504567
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04567
; FILING DATE: CONCURRENTLY HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,357
; FILING DATE: 13-APR-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: UOAB025P--
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4368 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 49..3123
;
PCT-US95-04567-3

```

```

Query Match 61.6%; Score 15.4; DB 5; Length 4368;
Best Local Similarity 76.0%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25
   ||| ||||| |||||
Db 3624 AATGAAGAAAGTCAAAAATATT 3648

RESULT 43
US-09-032-365A-11/c
; Sequence 11, Application US/09032365A
; Patent No. 6114502
; GENERAL INFORMATION:
; APPLICANT: No. 6114502th, Michael
; APPLICANT: Nishina, Patsy
; APPLICANT: Naggart, Juergen
; APPLICANT: No. 6114502en-Trauth, Konrad
; TITLE OF INVENTION: GENE FAMILY ASSOCIATED WITH
; TITLE OF INVENTION: NEUROSENSORY DEFECTS
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Avenue, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/032,365A
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: SEQ-2CIP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3400
; TELEFAX: 650-327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5994 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
;
US-09-032-365A-11

Query Match 61.6%; Score 15.4; DB 3; Length 5994;
Best Local Similarity 76.0%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25
   ||||| ||||| |||||
Db 3108 AGTGTCTCAAGTCCCGAGATATAATT 3084

RESULT 44
PCT-US95-11859-2/c
; Sequence 2, Application PC/TUS9511859
; GENERAL INFORMATION:

```

APPLICANT: JANSEN, KATHRIN U.  
APPLICANT: HOFMANN, KATHRYN J.  
TITLE OF INVENTION: DNA ENCODING HUMAN PAPILLOMA VIRUS TYPE  
TITLE OF INVENTION: 6A  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CHRISTINE E. CARTY  
STREET: 126 E. LINCOLN AVENUE - P.O. BOX 2000  
CITY: RAHWAY  
STATE: NJ USA  
COUNTRY: USA  
ZIP: 07065-0907  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/11859  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/310,468  
FILING DATE: 22-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: CARTY, CHRISTINE E.  
REGISTRATION NUMBER: 36,099  
REFERENCE/DOCKET NUMBER: 19307 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-6734  
TELEFAX: (908) 594-4720  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8010 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
PCT-US95-11859-2

Query Match 61.6%; Score 15.4; DB 5; Length 8010;  
Best Local Similarity 76.0%; Pred. No. 1.4e+02;  
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 attgtcgcaagtccacagaattattt 25  
||| |||| |||| |||| |||| ||  
Db 210 ATTCTGCAAAACACACAAATTAATT 186

RESULT 45  
US-08-952-127-11  
Sequence 11, Application US/08952127  
Patent No. 6211336  
GENERAL INFORMATION:  
APPLICANT: Shiloh, Yosef  
APPLICANT: Tagle, Danilo A.  
APPLICANT: Collins, Francis S.  
TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kohn & Associates  
STREET: 30500 No. 6211336thwestern Hwy., Suite 410  
CITY: Farmington Hills  
STATE: Michigan  
COUNTRY: U.S.  
ZIP: 48334  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/952,127  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Kohn, Kenneth I.  
REGISTRATION NUMBER: 30,995  
REFERENCE/DOCKET NUMBER: 2290.00029  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 810-539-5050  
TELEFAX: 810-539-5055  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9620 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
ORIGINAL SOURCE:  
ORGANISM: Mus musculus  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: Chromosome 9, Band 9C  
US-08-952-127-11

Query Match 61.6%; Score 15.4; DB 4; Length 9620;  
Best Local Similarity 76.0%; Pred. No. 1.4e+02;  
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtccacagaattattt 25  
||| |||| |||| |||| |||| ||  
Db 1234 ATCTTCAGAACTCACAGAGTGATTT 1258

Search completed: October 9, 2001, 11:39:32  
Job time: 1887 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 13:46:27 ; Search time 5930.9 Seconds  
(without alignments)  
39,846 Million cell updates/sec

Title: US-09-396-196F-2  
Perfect score: 25  
Sequence: 1 attgtcgcaagtacacagaattattt 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST: \*  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est25.\*  
26: gb\_est26.\*  
27: gb\_est27.\*  
28: gb\_est28.\*  
29: gb\_est29.\*  
30: gb\_est30.\*  
31: gb\_est31.\*  
32: gb\_est32.\*  
33: gb\_est33.\*  
34: gb\_est34.\*  
35: gb\_est35.\*  
36: gb\_est36.\*  
37: gb\_est37.\*  
38: gb\_est38.\*  
39: gb\_est39.\*  
40: gb\_est40.\*  
41: gb\_est41.\*  
42: gb\_est42.\*  
43: gb\_est43.\*  
44: gb\_est44.\*  
45: gb\_est45.\*  
46: gb\_est46.\*  
47: gb\_est47.\*

44: em\_esthum10.\*  
45: em\_esthum11.\*  
46: em\_esthum12.\*  
47: em\_esthum13.\*  
48: em\_esthum14.\*  
49: em\_esthum15.\*  
50: em\_esthum16.\*  
51: em\_esthum17.\*  
52: em\_esthum18.\*  
53: em\_esthum19.\*  
54: em\_esthum20.\*  
55: em\_esthum21.\*  
56: em\_esthum22.\*  
57: em\_esthum23.\*  
58: em\_esthum24.\*  
59: em\_esthum25.\*  
60: em\_esthum26.\*  
61: em\_esthum27.\*  
62: em\_esthum28.\*  
63: em\_estin1.\*  
64: em\_estin2.\*  
65: em\_estin3.\*  
66: em\_estin4.\*  
67: em\_estin5.\*  
68: em\_estom1.\*  
69: em\_estom2.\*  
70: em\_estov1.\*  
71: em\_estov2.\*  
72: em\_estpl1.\*  
73: em\_estpl2.\*  
74: em\_estpl3.\*  
75: em\_estpl4.\*  
76: em\_estpl5.\*  
77: em\_estpl6.\*  
78: em\_estpl7.\*  
79: em\_estpl8.\*  
80: em\_estpl9.\*  
81: em\_estpl10.\*  
82: em\_estrol.\*  
83: em\_estro2.\*  
84: em\_estro3.\*  
85: em\_estro4.\*  
86: em\_estro5.\*  
87: em\_estro6.\*  
88: em\_estro7.\*  
89: em\_estro8.\*  
90: em\_estro9.\*  
91: em\_estrol0.\*  
92: em\_estrol1.\*  
93: em\_estrol2.\*  
94: em\_estrol3.\*  
95: em\_estrol4.\*  
96: em\_estrol5.\*  
97: em\_estrol6.\*  
98: em\_estrol7.\*  
99: em\_estrol8.\*  
100: em\_estrol9.\*  
101: em\_estro20.\*  
102: gb\_est25.\*  
103: gb\_est26.\*  
104: gb\_est27.\*  
105: gb\_est28.\*  
106: gb\_est29.\*  
107: gb\_est30.\*  
108: gb\_est31.\*  
109: gb\_est32.\*  
110: gb\_est33.\*  
111: gb\_est34.\*  
112: gb\_est35.\*  
113: gb\_est36.\*  
114: gb\_est37.\*  
115: gb\_est38.\*  
116: gb\_est39.\*

```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	19.2	75.8	866	136	BE540498	BE540498 601066628
C 2	18.7	75.2	547	235	AQ897324	AQ897324 HS_3120_B
C 3	18.6	74.4	239	156	D21670	D21670 MUS81D04 mo
C 4	18.6	74.4	357	140	BE800706	BE800706 sq97c06.y
C 5	18.6	74.4	382	31	AV525093	AV525093 AV525093
C 6	18.6	74.4	388	120	AW781150	AW781150 sl89q06.y
C 7	18.6	74.4	444	189	T83852	T83852 yd66b07.sl
C 8	18.6	74.4	454	159	N58785	N58785 v76c09.sl
C 9	18.6	74.4	459	168	BF711742	BF711742 MI-P-E6-a
C 10	18.6	74.4	556	240	AZ277120	AZ277120 RPCI-23-1
C 11	18.6	74.4	563	239	AZ207443	AZ207443 SP_0135_A
C 12	18.6	74.4	576	233	AQ771996	AQ771996 HS_5413_B
C 13	18.6	74.4	596	238	AZ083812	AZ083812 RPCI-23-2
C 14	18.6	74.4	677	247	AZ650467	AZ650467 LM0520307
C 15	18.6	74.4	709	223	AQ020375	AQ020375 CIT-HSP-2
C 16	18.6	74.4	751	138	BE588717	BE588717 GM700007A
C 17	18.6	74.4	751	167	BE394042	BE394042 601312405
C 18	18.6	74.4	752	239	AZ183578	AZ183578 SP_1001_A
C 19	18.6	74.4	849	137	BE548526	BE548526 601075921
C 20	18.6	74.4	856	233	AQ740253	AQ740253 HS_5505_A
C 21	18.6	74.4	972	220	CNS02200	AL221439 Tetraodon
C 22	18.6	74.4	1015	171	BF966737	BF966737 602286714
C 23	18.2	72.8	169	190	W18068	W18068 mb85D03.r1
C 24	18.2	72.8	379	13	AA920661	AA920661 vy48c01.r
C 25	18.2	72.8	379	187	R21156	R21156 yb20a04.r1
C 26	18.2	72.8	441	235	AQ885836	AQ885836 HS_5527_B
C 27	18.2	72.8	632	229	AQ510232	AQ510232 nbxb0095A
C 28	18.2	72.8	666	225	AQ200982	AQ200982 RPCI11-61
C 29	18.2	72.8	965	221	CNS04111	AL270334 Tetraodon
C 30	17.8	71.2	289	223	AQ070825	AQ070825 HS_2256_A
C 31	17.8	71.2	557	233	AQ754255	AQ754255 HS_5326_B
C 32	17.8	71.2	646	148	BF404390	BF404390 UI-R-CA1-
C 33	17.6	70.4	163	13	AA927716	AA927716 om72f05.s
C 34	17.6	70.4	276	158	H60887	H60887 y46f09.r1
C 35	17.6	70.4	300	156	C34741	C34741 C34741 yu11
C 36	17.6	70.4	317	244	AZ456884	AZ456884 LM0260B02
C 37	17.6	70.4	329	21	AI495270	AI495270 sb02q09.y
C 38	17.6	70.4	335	158	H75480	H75480 yu61d12.sl
C 39	17.6	70.4	335	191	Z30954	Z30954 ATTS2670 ve
C 40	17.6	70.4	335	241	AZ289215	AZ289215 RPCI-23-1
C 41	17.6	70.4	359	223	AQ019978	AQ019978 CIT-HSP-2
C 42	17.6	70.4	416	149	BF509677	BF509677 UI-H-B14-
C 43	17.6	70.4	420	223	AQ008647	AQ008647 CIT-HSP-2
C 44	17.6	70.4	432	159	N20061	N20061 yx28f07.sl
C 45	17.6	70.4	445	229	AQ463730	AQ463730 HS_5051_B

## ALIGNMENTS

```

RESULT 1
BE540498/c
LOCUS BE540498 866 bp mRNA EST 09-AUG-2000
DEFINITION 601066628f1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3452959 5',
mRNA sequence.
ACCESSION BE540498
VERSION BE540498.1 GI:9769143
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 866)
NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT

```

```

Email: c9apbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM8435 row: g column: 08
High quality sequence stop: 593.
FEATURES
Location/Qualifiers
1..866
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3452959"
/clone_lib="NIH_MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.5 kb. Library prepared by Life
Technologies."
BASE COUNT 255 a 134 c 182 g 294 t 1 others
ORIGIN
Query Match 76.8%; Score 19.2; DB 136; Length 866;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 ttctgcagctacagaaattttt 25
Db 642 TTGTCCCCAGTCACACAAATTATTT 619
RESULT 2
AQ897324/c
LOCUS AQ897324 547 bp DNA GSS 10-NOV-1999
DEFINITION HS_3120_B2_B04_T7C CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3120 Col=8 Row=D, DNA sequence.
ACCESSION AQ897324
VERSION AQ897324.1 GI:6353514
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 547)
REFERENCE 1 (bases 1 to 547)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3120 row: D column: 8
Seq primer: T7
Class: BAC ends
High quality sequence stop: 547.
Location/Qualifiers
1..547
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3120 Col=8 Row=D"
FEATURES
source

```

```

/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/Note="Organ: sperm; Vector: pBel0BAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 127 a 111 c 94 g 210 t 5 others
ORIGIN

Query Match 75.2%; Score 18.8; DB 235; Length 547;
Best Local Similarity 90.9%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 tttgcgaagtcacagaattatt 24
||||| ||| ||| ||| |||
DB 354 TGTGCGAAGACACATAATTATT 333

RESULT 3
D21670/c
LOCUS D21670 239 bp mRNA EST 07-OCT-1996
DEFINITION M0581D04 mouse embryonal carcinoma cell line F9 Mus musculus cDNA
clone 81D04, mRNA sequence.
ACCESSION D21670
VERSION D21670.1 GI:618798
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 239)
AUTHORS Nishiguchi,S., Joh,T., Horie,K., Zou,Z., Yasunaga,T. and Shimada,K.
TITLE A survey of genes expressed in undifferentiated mouse embryonal
carcinoma F9 cells: Characterization of low-abundance mRNAs
J. Biochem. 116, 128-139 (1994)
JOURNAL 95096008
MEDLINE
COMMENT Contact: Kazunori Shimada
Department of Medical Genetics, Division of Molecular Biomedicine
Research Institute for Microbial Diseases, Osaka University
3-1, Yamadaoka, Suita, Osaka, 565, Japan
Tel: 06-879-8325
Fax: 06-879-8326.
FEATURES
Location/Qualifiers
source
1..239
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="81D04"
/clone_lib="mouse embryonal carcinoma cell line F9"
BASE COUNT 52 a 46 c 64 g 77 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 156; Length 239;
Best Local Similarity 84.0%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattatt 25
||||| ||| ||| ||| |||
DB 30 ATTGTCCCAATTACAGAACTATT 6

RESULT 4
BE800706
LOCUS BE800706 357 bp mRNA EST 20-SEP-2000
DEFINITION sq97c06.yl Gm-c1049 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-c1049-1091 5' similar to TR:082150 082150 CHLOROPLAST FTSH
PROTEASE. ;, mRNA sequence.
ACCESSION BE800706
VERSION BE800706.1 GI:10231818
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

```

```

Rosidae; eucosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
1 (bases 1 to 357)
Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Corvett,V., Khanna
,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Putative full length read
vector to vector length is 364
High quality sequence stop: 340.
FEATURES
Location/Qualifiers
source
1..357
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1049-1091"
/clone_lib="Gm-c1049"
/tissue_type="whole seedlings of greenhouse grown plants"
/dev_stage="3 week old"
/lab_host="DH10B"
/Note="Vector: pBluescript II SK+; Site.1: EcoRI; Site.2:
XhoI; The Clark NIL was constructed and seed was provided
by Dr. J. Specht, University of Nebraska (Shoemaker and
Specht, 1995). The cDNA library was constructed from mRNA
isolated from whole seedlings of 3 week old greenhouse
grown plants. Complementary DNA was synthesized from mRNA
using a primer consisting of a poly(dT) sequence with a
XhoI restriction site and a 3' anchor. EcoRI adapters were
ligated to the blunt-ended cDNA fragments followed by XhoI
digestion. The cDNA fragments were directionally cloned
into the EcoRI-XhoI restriction site of the pBluescript
vector. The ligated cDNA fragments were transformed into
DH10B host cells (GibcoBRL). The library was constructed
in cooperation with Dr. Paul Keim's laboratory at Northern
Arizona University."
BASE COUNT 169 a 53 c 47 g 88 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 140; Length 357;
Best Local Similarity 84.0%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattatt 25
||||| ||| ||| ||| |||
DB 133 ATACACGCAAGTCACAGATTATGT 157

RESULT 5
AV525093
LOCUS AV525093 382 bp mRNA EST 01-SEP-2000
DEFINITION AV525093 Arabidopsis thaliana aboveground organs two to six-week
old Arabidopsis thaliana cDNA clone APD17c09R 5', mRNA sequence.
ACCESSION AV525093
VERSION AV525093.1 GI:8684621
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 382)  
Asamizu,E., Nakanura,Y., Sato,S. and Tabata,S.  
A large scale analysis of cDNA in Arabidopsis thaliana: Generation  
of 12,028 non-redundant expressed sequence tags from normalized and  
size-selected cDNA libraries  
DNA Res. 7, 175-180 (2000)  
20363093  
Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
source  
1..382  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="AP017c09R"  
/clone\_lib="Arabidopsis thaliana aboveground organs two to  
six-week old"  
/tissue\_type="aboveground organs"  
/dev\_stage="two to six-week old"  
/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
XhoI"

BASE COUNT 101 a 87 c 89 g 105 t  
ORIGIN  
1 attgtcgcaagtccacagaattattt 25  
||||| ||||||| ||||| |||||  
Db 296 APTGTTGCAAGTCAGAGAAGTCTTT 320

Query Match 74.4%; Score 18.6; DB 31; Length 382;  
Best Local Similarity 84.0%; Pred. No. 2e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 attgtcgcaagtccacagaattattt 25  
||||| ||||||| ||||| |||||  
Db 296 APTGTTGCAAGTCAGAGAAGTCTTT 320

RESULT 6  
AW781150 388 bp mRNA EST 12-MAY-2000  
LOCUS  
DEFINITION  
s189906.y1 Gm-cl037 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
Gm-cl037-1235 5' similar to TR:082150 082150 CHLOROPLAST FTSH  
PROTEASE. ; mRNA sequence.  
AW781150  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
soybean.  
Glycine max  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycyne.  
1 (bases 1 to 388)  
Shoemaker,R., Keim,P., Vodkin,L., Erpelidng,J., Coryelli,V., Khanna  
,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,  
Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers  
,Y., Person,B., Swaller,T., Gibbons,M., Pepe,D., Harvey,N., Schurk  
,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann  
,R., Waterston,R. and Wilson,R.  
Public Soybean EST Project  
Unpublished (1999)  
Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project

TITLE  
JOURNAL  
COMMENT  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available through: Genome Systems, Inc. 4633 World  
Parkway Circle St. Louis, Missouri 63134 For further information  
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)  
427-3324 or contact: clones@genomesystems.com or

info@genomesystems.com web site: www.genomesystems.com  
High quality sequence stop: 375.

FEATURES  
source  
1..388  
Location/Qualifiers  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-cl037-1235"  
/clone\_lib="Gm-cl037"  
/tissue\_type="fully expanded leaves of greenhouse grown  
plants"  
/dev\_stage="2 week old"  
/lab\_host="DH10B"  
/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; This  
cDNA library was constructed from mRNA isolated from fully  
expanded leaves of greenhouse grown plants that were 2  
weeks old. The library was prepared using the Life  
Technologies pSuperScript cDNA library construction kit.  
Complementary DNA was synthesized from mRNA using a  
poly(dT) sequence with a NotI restriction site. SalI  
linkers adapters were ligated to the blunt-ended cDNA  
fragments followed by NotI digestion. The cDNA fragments  
were directionally cloned into the NotI-SalI restriction  
site of the pSPORT1 vector. The ligated cDNA fragments  
were transformed into E.coli Electro-Max DH10B host  
cells. This library was constructed in the laboratory of  
Dr. Lila Vodkin by Anu Khanna at the University of  
Illinois at Urbana- Champaign. email: l-vodkin@uiuc.edu"

BASE COUNT 123 a 67 c 92 g 106 t  
ORIGIN  
1 attgtcgcaagtccacagaattattt 25  
||||| ||||||| ||||| |||||  
Db 329 ATACCGCAAGTCACAGATTATGT 353

Query Match 74.4%; Score 18.6; DB 120; Length 388;  
Best Local Similarity 84.0%; Pred. No. 2e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 attgtcgcaagtccacagaattattt 25  
||||| ||||||| ||||| |||||  
Db 329 ATACCGCAAGTCACAGATTATGT 353

RESULT 7  
T83852 444 bp mRNA EST 16-MAR-1995  
LOCUS  
DEFINITION  
Yd66b07.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone  
IMAGE:113173 3' similar to gb:1M87934 HUMANAL043 Human carcinoma  
cell-derived Alu RNA transcript, (rRNA); gb:X54150.rnal  
IMMUNOGLOBULIN ALPHA FC RECEPTOR PRECURSOR (HUMAN); contains Alu  
repetitive element; contains L1 repetitive element ; mRNA sequence.  
T83852  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 444)  
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman  
,M., Hultman,M., Kucaba,T., Le.M., Lennon,G., Marra,M., Parsons,J.,  
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston  
,R., Williamson,A., Wohldmann,P. and Wilson,R.  
The WashU-Merck EST Project  
Unpublished (1995)  
Contact: Wilson RK

TITLE  
JOURNAL  
COMMENT  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 939  
High quality sequence stops: 334 Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 939 Std Error: 0.00



polylinker; Site1: Not I; Site2: EcoRI; The MI-P-E6 library is derived from fetus at gestational day 20. For a detailed description of the library from which this clone was derived, please visit our web site at <http://pigest.genome.iasstate.edu/>.

TAG\_LIB-MI-P-E6  
TAG\_TISSUE=fetus at gestational day 20  
TAG\_SEQ=AGGAA

BASE COUNT 157 a 91 c 81 g 130 t  
ORIGIN

Query Match 74.4%; Score 18.6; DB 168; Length 459;  
Best Local Similarity 84.0%; Pred. No. 2e+02; Indels 0; Gaps 0;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25  
||||| ||||| ||||| ||||| |||||  
Db 385 ATTGTGCAATTCACAGAGTACTTT 361

RESULT 10  
AZ277120/c  
LOCUS  
DEFINITION RPCI-23-134L7\_TV RPCI-23 Mus musculus genomic clone RPCI-23-134L7, DNA sequence. GSS 26-JUL-2000

ACCESSION AZ277120 GI:9496111

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Other GSSs: RPCI-23-134L7.TJ  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 134 row: L column: 7  
Seq primer: T7  
Class: BAC ends.

FEATURES  
source  
1. .556  
Location/Qualifiers

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-134L7"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site1: EcoRI; Site2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."  
181 a 109 c 131 g 135 t

BASE COUNT

ORIGIN

Query Match 74.4%; Score 18.6; DB 240; Length 556;  
Best Local Similarity 84.0%; Pred. No. 2e+02; Indels 0; Gaps 0;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25  
||||| ||||| ||||| ||||| |||||  
Db 156 ATTGTGCAATTCACAGAGTACTTT 132

RESULT 11  
AZ207443/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Additional resources

Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)

20402566

Contact: Cameron, RA, Davidson, EH, Hood, L

Division of Biology 156-29

California Institute of Technology

Pasadena California 91125, USA

Tel: (626) 395-8421

Fax: (626) 793-3047

Email: acameron@caltech.edu

Plate: 135 row: G column: 7

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 563.

Location/Qualifiers

1. .563

/organism="Strongylocentrotus purpuratus"

/db\_xref="taxon:7668"

/clone="Plate-135 Col=7 Row=G"

/clone\_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"

/note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli DH10B"

BASE COUNT 198 a 84 c 132 g 149 t

ORIGIN

Query Match 74.4%; Score 18.6; DB 239; Length 563;

Best Local Similarity 84.0%; Pred. No. 2e+02; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25

||||| ||||| ||||| ||||| |||||

Db 89 ACTGTCAAGTCTCAGATCATTT 65

RESULT 12

AQ771996/c

LOCUS

DEFINITION

Human Male BAC Library Homo sapiens genomic clone Plate=989 Col=14 Row=L, DNA sequence.

576 bp DNA GSS 29-JUL-1999



Fax: 801 585 7177  
Email: ddunne@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0520 row: J column: 07  
Seq primer: CACACAGGAACACGTATGACC  
Class: plasmid ends  
High quality sequence stop: 677.

## FEATURES

```

1. .677
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0520J07"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (g14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
210 a 168 c 154 g 145 t

```

Fax: 301 838 0208  
Email: [mdadams@tigr.org](mailto:mdadams@tigr.org)  
Clones are available from Research Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC  
end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html).  
Seq primer: M13 Reverse  
Class: BAC ends.

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

```

source
1. .709
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2301C14"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/notes="Vector: pBelOBAC11; Site_1: HindIII; Site_2: HindIII"
BASE COUNT      303 a 109 c 124 g 173 t
ORIGIN
Query Match      74.4%; Score 18.6; DB 223; Length 709;
Best Local Similarity 84.0%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattattt 25
||||| ||||| ||||| ||||| |||
Db 664 ATTGTTGCAAGTAAACAGAAATTTCT 640

RESULT 16
BE58717/c
LOCUS      BE58717/c              709 bp      mRNA      EST      06-SEP-2000
DEFINITION GM700007A10F9 Gm-r1070 Glycine max cDNA clone Gm-r1070-2561 3',
mRNA sequence.
ACCESSION  BE58717
VERSION    BE58717
KEYWORDS   BE58717.1 GI:9984609
SOURCE     soybean.
ORGANISM   Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE  1 (bases 1 to 751)
AUTHORS    Vodkin,L., Kelm,P., Shoemaker,R., Retzel,E., Khanna,A., Corvelli,V.,
Erpelding,J., Raph,C., Shoop,E., Pardini,J., Liu,L. and Lewin,H.
A Functional Genomics Program for Soybean (NSF 9872565)
Unpublished (1999)
OTHER_ESTS AW156684 corresponding to Gm-cl015-2632 (5')
Contact: Vodkin, L.O., PI, A Functional Genomics Program for
Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional
Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134. For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or info@genome
systems.com web site: www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.

```

FEATURES  
SOURCE

```

source
1. .751
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="Gm-r1070-2561"
/clone_lib="Gm-r1070"
/note="The library Gm-r1070 is a sequence-driven, rereacted
set of 9,216 clones selected from cDNA libraries from
various tissues and stages of development of soybean that

```





```

RESULT 19
BE548526/c
LOCUS BE548526 849 bp mRNA EST 09-AUG-2000
DEFINITION 601075921F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3461861 5',
mRNA sequence.
ACCESSION BE548526
VERSION BE548526.1 GI:9777171
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 849)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1A8458 row: j column: 06
High quality sequence stop: 568.
Location/Qualifiers
1..849
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3461861"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."
BASE COUNT 248 a 226 c 159 g 216 t
ORIGIN
source
Query Match 74.4%; Score 18.6; DB 137; Length 849;
Best Local Similarity 84.0%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtcacagaattattt 25
||||| ||||| ||||| ||||| |||||
Db 557 ACTGTCACAAAGTCACAGATTATGT 533

RESULT 20
A0740253/c
LOCUS A0740253 856 bp DNA GSS 16-JUL-1999
DEFINITION HS_5505_A2_F08_37A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=1081 Col=16 Row=K, DNA sequence.
ACCESSION A0740253
VERSION A0740253.1 GI:5517775
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 856)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L

```

```

High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1081 row: K column: 16
Seq primer: T7
Class: BAC ends
High quality sequence stop: 856.
Location/Qualifiers
1..856
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1081 Col=16 Row=K"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"
BASE COUNT 282 a 142 c 176 g 254 t
ORIGIN
source
Query Match 74.4%; Score 18.6; DB 233; Length 856;
Best Local Similarity 84.0%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtcacagaattattt 25
||||| ||||| ||||| ||||| |||||
Db 552 ATTGTCACAAATACAGATTCTTT 528

RESULT 21
CNS022UU/c
LOCUS CNS022UU 972 bp DNA GSS 15-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
183108 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL221439,1 GI:7880258
VERSION AL221439
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 972)
AUTHORS Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 972)
AUTHORS Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 972)
AUTHORS Genoscope.
TITLE Direct Submission

```

JOURNAL  
COMMENT

Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/tetraodon>.

## FEATURES

source

1.972  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COAG183BE04SP1-end ;  
PUC-Ori"

BASE COUNT 262 a 196 c 218 g 266 t 30 others  
ORIGIN

## Query Match

Best Local Similarity 74.4%; Score 18.6; DB 220; Length 972;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtcacagaattattt 25  
||||| ||||| ||||| ||||| |||||

Db 164 ATTTAGCAATTCACAGATTATTAT 140

## RESULT 22

BF966737 1015 bp mRNA EST 23-JAN-2001  
LOCUS 60228671AT1 NIH\_MGC\_95 Homo sapiens cDNA clone IMAGE:4375802 3',  
DEFINITION mRNA sequence.

ACCESSION BF966737

VERSION BF966737.1 GI:12333952

KEYWORDS EST.

SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1015)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [ccapbs-remail.nih.gov](mailto:ccapbs-remail.nih.gov)

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

Toshiyuki and Piero Carninci (RIKEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L1AM10041 row: k column: 03

High quality sequence start: 25

High quality sequence stop: 735.

## FEATURES

source

1.1015  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4375802"  
/clone\_lib="NIH\_MGC\_95"  
/tissue\_type="hippocampus"  
/lab\_host="DH108"

/note="Organ: brain; Vector: pBluescriptR (modified  
pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag  
); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTTWN-3',  
size-selected for average insert size 2.5 kb and  
normalized to ROT 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NIH/NHGRI, National  
Institutes of Health). Note: this is a NIH\_MGC Library."

## BASE COUNT

273 a 218 c 236 g 288 t

## ORIGIN

Query Match 74.4%; Score 18.6; DB 171; Length 1015;  
Best Local Similarity 84.0%; Pred. No. 2.2e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtcacagaattattt 25

||||| ||||| ||||| ||||| |||||

Db 455 AATGGCGCTAGTCACAGATTATTAT 479

## RESULT 23

W18068

LOCUS W18068

DEFINITION

W18068 169 bp mRNA EST 10-SEP-1996  
mb85d03.r1 Soares mouse p3NMFL19.5 Mus musculus cDNA clone  
IMAGE:336197 5' similar to PIR:A48080 A48080 basic helix-loop-helix  
protein TFEC - rat ; mRNA sequence.

ACCESSION W18068

VERSION W18068.1 GI:1292539

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 169)

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and

Waterston, R.

The WashU-HHMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)

This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

MGI:217597

Possible reversed clone: similarity on wrong strand

Seq primer: mob.REGA+ET

High quality sequence stop: 142.

## FEATURES

source

1.169  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:336197"  
/dev\_stage="19.5 dpc total fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pT73D (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I oligo(dT) primer [5',  
TCTTACCAATCTGAAGTGGAGCGCCGCAATTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaldo. RNA was kindly provided by  
Dr. Minoru Ko (Wayne State University)."

## BASE COUNT

39 a 45 c 36 g 49 t

## Query Match

Best Local Similarity 72.8%; Score 18.2; DB 190; Length 169;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattatt 23  
||||| || ||||||| |||||  
Db 40 ATTGCTCACGCACAGATTAT 62

## RESULT 24

AA920661 379 bp mRNA EST 20-APR-1998  
LOCUS  
DEFINITION VY48c01.r1 Stragatene mouse lung 937302 Mus musculus cDNA clone  
IMAGE:1298688 5' similar to WP:F43E2.7 CE10348 ; mRNA sequence.  
ACCESSION AA920661  
VERSION  
KEYWORDS  
SOURCE EST.  
ORGANISM house mouse.  
Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS 1 (bases 1 to 379)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:679736

Putative full length read

vector to vector length is 398

Seq primer: -28ml3 revl Et from Amersham

High quality sequence stop: 370.

## FEATURES

source

1. .379  
/organism="Mus musculus"  
/strain="C57BL/6 x CBA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1298688"  
/clone\_lib="Stragatene mouse lung 937302"  
/sex="female"  
/tissue\_type="lung"  
/dev\_stage="6-8 month old"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: lung; Vector: pBluescript SK-; Site\_1: EcoRI  
; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo  
dn. 6-8 month old female lung and 1.5 year old male lung  
were source of mRNA. Average insert size: 1.5 kb; Uni-ZAP  
XR vector: -5' adaptor sequence: 5' GAATTCGCGCAGAG 3' -3'  
adaptor sequence: 5' CTCAGCTTTTTTTTTTTTTTTT 3'"

BASE COUNT 102 a 69 c 87 g 121 t  
ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 379;  
Best Local Similarity 87.0%; Pred. No. 3e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 attgtcgcaagtcacagaattatt 23  
||||| || ||||||| |||||  
Db 303 ATAGTCGAATGCACAGATTAT 325

## RESULT 25

R21156/c 379 bp mRNA EST 18-APR-1995  
LOCUS  
DEFINITION yh20a04.r1 Soares placenta Nb2HP Homo sapiens cDNA clone

ACCESSION  
VERSION  
KEYWORDS  
SOURCE

ORGANISM

Homo sapiens

human.

IMAGE:130254 5', mRNA sequence.

R21156

R21156.1 GI:775937

EST.

human.

REFERENCE

1 (bases 1 to 379)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman

M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,

Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston

R., Williamson, A., Wohlmann, P. and Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 987

High quality sequence stops: 337

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 987 Std Error: 0.00

Seq primer: M13RP1

High quality sequence stop: 337.

Location/Qualifiers

1. .379

/organism="Homo sapiens"

/db\_xref="GDB:537603"

/db\_xref="taxon:9606"

/clone="IMAGE:130254"

/clone\_lib="Soares placenta Nb2HP"

/sex="Female"

/dev\_stage="placenta obtained at birth (full term)"

/lab\_host="DH10B (ampicillin resistant)"

/note="Organ: placenta; Vector: pT73D (Pharmacia) with a

modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'

AACTGGAAGAAATTCGCGCGCAGGAATTTTTTTTTTTT 3']

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization. Library

constructed by Bento Soares and M.Fatima Bonaldo.

BASE COUNT 123 a 56 c 83 g 117 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 187; Length 379;

Best Local Similarity 87.0%; Pred. No. 3e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattatt 24

||||| ||||||| |||||

Db 325 TTGTCCCAAGTCACACAAATTAGT 303

RESULT 26

AQ885836 441 bp DNA GSS 09-NOV-1999

LOCUS HS\_5527\_B2\_C10\_SP6E RPCI-11 Human Male BAC Library Homo sapiens

DEFINITION genomic clone Plate=9295 Col=20 Row=F, DNA sequence.

ACCESSION AQ885836

VERSION AQ885836.1 GI:6317303

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE  
AUTHORS 1 (bases 1 to 441)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallaceu.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@jeng.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm))  
or from Research Genetics (<http://www.htsc.washington.edu>)  
Plate: 9295 row: F column: 20  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 441.  
Location/Qualifiers  
1. 441  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate-9295 Col-20 Row-F"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBAC3.6 vector at EcoRI sites"  
BASE COUNT 116 a 66 c 81 g 175 t 3 others  
ORIGIN  
Query Match 72.8%; Score 18.2; DB 235; Length 441;  
Best Local Similarity 87.0%; Pred. No. 3e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 ttgtcgcaagtcacagaattatt 24  
|||||  
Db 198 TTGTCGCAAGTCTTAGAATTCTT 220  
|||||  
RESULT 27  
AQ510232 632 bp DNA GSS 04-MAY-1999  
LOCUS nbxb0095A23f CUGI Rice BAC Library Oryza sativa genomic clone  
DEFINITION nbxb0095A23f, DNA sequence.  
ACCESSION AQ510232  
VERSION AQ510232.1 GI:4732836  
KEYWORDS GSS.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.  
1 (bases 1 to 632)  
Wing,R.A. and Dean,R.A.  
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome  
JOURNAL Unpublished (1998)  
COMMENT Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293

Email: rwing@clemson.edu  
Seq primer: TAATAGCACTCACTATAGGG  
Class: BAC ends  
High quality sequence stop: 372.  
Location/Qualifiers  
1. 632  
/organism="Oryza sativa"  
/strain="Japonica"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/clone="nbxb0095A23f"  
/clone\_lib="CUGI Rice BAC Library"  
/tissue\_type="Leaf"  
/lab\_host="E. coli DH10B"  
/note="Vector: pBeloBAC11; Site\_1: HindIII; Site\_2:  
HindIII; Rice is one of two most popular grains in the  
world. Half of the world population especially those  
inhabiting highly populated areas of the humid tropics  
and subtropics, rely on rice as their primary source of  
carbohydrate. Monocotyledonous rice is a diploid plant  
(2n=24) with a haploid genome equivalent of 431 Mbp  
(Arumuganathan and Earle, 1991). The relatively small  
genome of rice, three times larger than that of  
Arabidopsis, makes it suitable for genomic studies. In  
order to facilitate positional cloning, physical mapping  
and genome sequencing of rice, we have constructed a BAC  
library from Oryza sativa, Nipponbare variety. The  
library contains 36,864 clones with an average insert size  
of 128.5 Kb providing 10.9 haploid genome equivalents. The  
deep coverage allows the isolation a particular sequence  
with a probability of 99.9%. Two high density filters,  
each containing 18,432 clones (doubly spotted), represent  
the whole library for colony screening."  
BASE COUNT 217 a 118 c 109 g 187 t 1 others  
ORIGIN  
Query Match 72.8%; Score 18.2; DB 229; Length 632;  
Best Local Similarity 87.0%; Pred. No. 3.2e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 3 ttgtcgcaagtcacagaattatt 25  
|||||  
Db 450 TGTCACAAGTCAATAATTATT 472  
|||||  
RESULT 28  
AQ200982 666 bp DNA GSS 20-APR-1999  
LOCUS RPCI11-61E22.TK RPCI-11 Homo sapiens genomic clone RPCI-11-61E22,  
DEFINITION DNA sequence.  
ACCESSION AQ200982  
VERSION AQ200982.1 GI:3613181  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 666)  
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,  
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.  
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building  
JOURNAL Unpublished (1998)  
COMMENT Other\_GSSs: RPCI11-61E22.TJ  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics ([inforesgen.com](http://inforesgen.com)). BAC end search page: [http://www.tigr.org/tldb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html)  
Class: BAC ends.

#### FEATURES

source  
1. .666  
/organism="Homo sapiens"  
/db\_xref="GDB:752315"  
/db\_xref="taxon:9606"  
/clone="RPC1-11-61E22"  
/clone\_lib="RPC1-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPC111 Human Male BAC Library"  
BASE COUNT 224 a 117 c 148 g 175 t 2 others  
ORIGIN  
Query Match 72.8%; Score 18.2; DB 225; Length 666;  
Best Local Similarity 87.0%; Pred. No. 3.2e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 tgtcgcaagtcacagaattattt 25  
||||| ||||| ||||| ||||| |||||  
Db 312 TGTGGAAAGTACAGATTATT 334

#### RESULT 29

CNS041L1/c 965 bp DNA GSS 18-MAY-2000  
LOCUS  
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone  
074L15 of library G from Tetraodon nigroviridis, genomic survey  
sequence.

ACCESSION AL270334  
VERSION AL270334.1 GI:7992248  
KEYWORDS GSS; genome survey sequence.  
SOURCE Tetraodon nigroviridis.  
ORGANISM Tetraodon nigroviridis.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 965)  
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,  
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 965)  
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.

TITLE Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence

JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 965)  
AUTHORS Genoscope.

TITLE Direct Submission  
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
COMMENT This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/tetraodon>.

#### FEATURES

source  
1. .965  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="074L15"  
/clone\_lib="G"  
/note="Genoscope sequence ID : C0BG074CF08LP1-end : T7"

BASE COUNT 285 a 170 c 193 g 316 t 1 others  
ORIGIN

Query Match 72.8%; Score 18.2; DB 221; Length 965;  
Best Local Similarity 87.0%; Pred. No. 3.3e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 tgtcgcaagtcacagaattattt 25  
||||| ||||| ||||| ||||| |||||  
Db 565 TGTCACTAGTCACTGAATTATT 543

#### RESULT 30

AQ070825/c 289 bp DNA GSS 05-AUG-1998  
LOCUS  
DEFINITION HS\_2256.AL\_F04\_MF CIT Approved Human Genomic Sperm Library D Homo  
sapiens genomic clone Plate=2256 Col=7 Row=K, DNA sequence.

ACCESSION AQ070825  
VERSION AQ070825.1 GI:3389984  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 289)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome

JOURNAL Proc.Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887

Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 2256 row: K column: 7

Class: BAC ends  
High quality sequence stop: 289.

FEATURES  
source  
1. .289  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="plate=2256 Col=7 Row=K"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in  
E-Coli DH10B"

BASE COUNT 117 a 76 c 39 g 57 t  
ORIGIN

Query Match 71.2%; Score 17.8; DB 223; Length 289;  
Best Local Similarity 90.5%; Pred. No. 4.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 tcgcaagtcacagaattattt 25  
||||| ||||| ||||| ||||| |||||  
Db 40 TCGCAAGTCACAGATTATT 20

#### RESULT 31

AQ754255/c 557 bp DNA GSS 27-JUL-1999  
LOCUS  
DEFINITION HS\_5326.B2\_C11\_T7A RPC1-11 Human Male BAC Library Homo sapiens  
genomic clone Plate=902 Col=22 Row=F, DNA sequence.

ACCESSION AQ754255  
VERSION AQ754255.1 GI:5618418

KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 557)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm))  
or from Resear h Genetics (<http://resgen.com>). BAC end Web Server:  
<http://www.htsc.washington.edu>  
Plate: 902 row: F column: 22  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 557.  
Location/Qualifiers  
1. .557  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="plate-902 Col=22 Row=F"  
/clone\_lib="RPCI-11 Human Male BAC Library"  
/sex="male"  
/note="vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBACE3.6 vector at EcoRI sites"  
BASE COUNT 148 a 109 c 144 g 152 t 4 others  
ORIGIN  
Query Match 71.2%; Score 17.8; DB 233; Length 557;  
Best Local Similarity 90.5%; Pred. No. 4.8e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 3 tgtcgaagtcacagaattat 23  
||||| ||| ||||| |||||  
Db 410 TGTCACTACACAGATTAT 390  
RESULT 32  
BF404390  
LOCUS BF404390 646 bp mRNA EST 28-NOV-2000  
DEFINITION UI-R-CAL-bie-f-08-0-UI.s1 UI-R-CAL Rattus norvegicus cDNA clone  
UI-R-CAL-bie-f-08-0-UI 3', mRNA sequence.  
ACCESSION BF404390  
VERSION BF404390.1 GI:11392365  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 646)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477  
COMMENT Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. the sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to identify it as a clone from the  
normalized corpus-striatum library cDNA Library Preparation: M.B.  
Soares Lab Clone distribution: clones will be available through  
Research Genetics ([www.resgen.com](http://www.resgen.com)) The following repetitive  
elements were found in this cDNA sequence: 382-596, >B3#SINE/B2  
Seq primer: M13 Forward  
POLYA=Yes.  
Location/Qualifiers  
1. .646  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-CAL-bie-f-08-0-UI"  
/clone\_lib="UI-R-CAL"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-CAL  
library is a subtracted library derived from the following  
tissues: thalamus, cerebellum, hypothalamus, medulla, pons  
, midbrain, cerebral cortex, corpus striatum, testis, and  
hippocampus. For a detailed description of the library  
from which this clone was derived, please visit our web  
site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has been  
previously described in (Bonaldo, Lennon and Soares,  
Genome Research 6:791-806, 1996)  
TAG\_LIB=UI-R-CAL  
TAG\_TISSUE=corpus-striatum  
TAG\_SEQ=CTAGG"  
BASE COUNT 159 a 117 c 132 g 238 t  
ORIGIN  
Query Match 71.2%; Score 17.8; DB 148; Length 646;  
Best Local Similarity 90.5%; Pred. No. 4.9e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 5 tcgaagtcacagaattattt 25  
| ||||| ||||| ||||| |||||  
Db 160 TTGCAAGACACAGAAATTATT 180  
RESULT 33  
AA927716  
LOCUS AA927716 163 bp mRNA EST 17-JUN-1998  
DEFINITION om72f05.s1 NCI-CGAP-GC4 Homo sapiens cDNA clone IMAGE:1552737 3',  
mRNA sequence.  
ACCESSION AA927716  
VERSION AA927716.1 GI:3076536  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 163)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael









JOURNAL  
COMMENT

Unpublished (1999)  
Other\_GSSs: RPCI-23-144014.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics ([info@resgen.com](http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html)). BAC end page: [http://www.tigr.org/tldb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html)  
plate: 144 row: 0 column: 14  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
1. .335  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-144014"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."  
BASE COUNT 102 a 55 c 52 g 126 t  
ORIGIN

Query Match 70.4%; Score 17.6; DB 241; Length 335;  
Best Local Similarity 83.3%; Pred. No. 5.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 attgtcgcaagtcacagaattatt 24  
||||| ||||||| ||| ||  
Db 247 ATTGTCACAGTCACAAATAGTT 224

## RESULT 41

AQ019978  
LOCUS  
DEFINITION C17-HSP-2302N11.TR CIT-HSP Homo sapiens genomic clone 2302N11, DNA sequence.  
ACCESSION AQ019978  
VERSION AQ019978.1 GI:3198714  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 359)  
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.  
Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)

## JOURNAL

COMMENT  
Other\_GSSs: CIT-HSP-2302N11.TF  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

Email: mdamas@tigr.org  
Clones are available from Research Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC end search page: [http://www.tigr.org/tldb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html).  
Seq primer: M13 Reverse  
Class: BAC ends.  
FEATURES  
source

Location/Qualifiers  
1. .359  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2302N11"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelOBAC11; Site\_1: HindIII; Site\_2: HindIII"  
BASE COUNT 92 a 85 c 60 g 122 t  
ORIGIN

Query Match 70.4%; Score 17.6; DB 223; Length 359;  
Best Local Similarity 83.3%; Pred. No. 5.7e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcacagaattatt 25  
||||| ||||||| |||||||  
Db 203 TTATCCAAAGTCACAGTAATTATTT 226

## RESULT 42

BF509677  
LOCUS  
DEFINITION UI-H-B14-apb-d-06-0-UI.s1 NCI\_CGAP\_Sub8 Homo sapiens cDNA clone IMAGE:3086651 3', mRNA sequence.  
ACCESSION BF509677.1 GI:11592975  
VERSION BF509677  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 416)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: [www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)  
Seq primer: M13 Forward  
POLYA=Yes.  
FEATURES  
source

Location/Qualifiers  
1. .416  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3086651"  
/clone\_lib="NCI\_CGAP\_Sub8"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; NCI\_CGAP\_Sub8 is a subtrated library derived from NCI\_CGAP\_Sub5. The NCI\_CGAP\_Sub8 library had 2.5 million recombinants. A single-stranded DNA preparation of NCI\_CGAP\_Sub5 was used as a tracer in a subtractive hybridization with a driver comprising: a pool of clones from NCI\_CGAP\_Sub5 (IMAGE clone ids 2732833-2737415, 3068040-3069191; 25% of the driver population), a pool of clones from NCI\_CGAP\_Sub4

(IMAGE clone ids 2723592-2729326; 25% of the driver population). NCI\_CGAP\_Sub6 (pool AIF-AJU, IMAGE ids 2728969-2733190; 25% of the driver population), and NCI\_CGAP\_Sub7 (IMAGE ids 3069192-3072238, 3081864-3084550; 25% of the driver population). Subtraction was performed as previously described (Bonald, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806.  
TAG\_SEQ=None found"

BASE COUNT 143 a 68 c 76 g 129 t  
ORIGIN

Query Match 70.4%; Score 17.6; DB 149; Length 416;  
Best Local Similarity 83.3%; Pred. No. 5.8e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttgtcgcaagtcacagaattattt 25

Db 14 TTTTCTCAAGTCACATATTATT 37

RESULT 43

AQ008647 420 bp DNA GSS 27-JUN-1998  
LOCUS CIT-HSP-2283J6.TRB CIT-HSP Homo sapiens genomic clone 2283J6, DNA  
DEFINITION sequence.

ACCESSION AQ008647

VERSION AQ008647

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 420)

AUTHORS Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wibie, C., Shizuaya, H., Simon, M., and Venter, J.C.

TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)

JOURNAL Unpublished (1998)

COMMENT Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC end search page:

http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html

Seq primer: M13 Reverse

Class: BAC ends.

FEATURES Location/Qualifiers

1..420

/organism="Homo sapiens"

/db\_xref="GDB:7148025"

/db\_xref="taxon:9606"

/clone="2283J6"

/clone\_lib="CIT-HSP"

/sex="Male"

/cell\_type="Sperm"

/note="vector: pBelOBAC11; Site\_1: HindIII; Site\_2:

HindIII"

BASE COUNT 117 a 79 c 90 g 134 t

ORIGIN

Query Match

Best Local Similarity 70.4%; Score 17.6; DB 223; Length 420;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttgtcgcaagtcacagaattattt 25

Db 227 TTGTTGCAATAACAGATTTTT 250

RESULT 44

N20061

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 432)

AUTHORS

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlffing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

High quality sequence stops: 365

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1148 Std Error: 0.00

Seq primer: ml3 -40 forward

High quality sequence stop: 365.

Location/Qualifiers

1..432

/organism="Homo sapiens"

/db\_xref="GDB:3872719"

/db\_xref="taxon:9606"

/clone="IMAGE:263077"

/clone\_lib="Soares melanocyte 2NbHM"

/sex="Male"

/tissue\_type="melanocyte"

/lab\_host="DH10B (ampicillin resistant)"

/note="vector: pT7T3D (Pharmacia) with a modified

polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGAGCGCGAGTTTTTTTTTTTTTTT 3'],

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT7T3 vector

(Pharmacia). Library constructed by Bento Soares and

M. Fatima Bonaldo. RNA from normal foreskin melanocytes

(FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 152 a 70 c 78 g 127 t

ORIGIN

Query Match

Best Local Similarity 70.4%; Score 17.6; DB 159; Length 432;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttgtcgcaagtcacagaattattt 25

Db 5 TTTTCTCAAGTCACATATTATT 28

RESULT 45

AQ463730/c

LOCUS

AQ463730

445 bp DNA

GSS

23-APR-1999

DEFINITION HS\_5051\_B2\_C10\_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=627 Col=20 Row=F, DNA sequence.

ACCESSION AQ463730

VERSION AQ463730.1 GI:4640825

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 445)

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering.bac.htm>) or from Resear h Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>  
Plate: 627 row: F column: 20  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 445.

FEATURES

source

1..445

Location/Qualifiers

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate=627 Col=20 Row=F"

/clone\_lib="RPCI-11 Human Male BAC Library"

/sex="male"

/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"

BASE COUNT 176 a 79 c 82 g 108 t

ORIGIN

Query Match 70.4%; Score 17.6; DB 229; Length 445;

Best Local Similarity 83.3%; Pred. No. 5.8e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgtcqaagtcacagaattattt 25  
||||| ||| || ||||| |||||

Db 202 TTGTAGCATGTCACAGAATTTT 179

Search completed: October 9, 2001, 13:46:30  
Job time: 9505 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 12:09:34 ; Search time 1666.31 Seconds  
(without alignments)  
232.066 Million cell updates/sec

Title: US-09-396-196F-3  
Perfect score: 25  
Sequence: 1 gtgtcgtgattcgtctgtttgaagcg 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

##### GenEmbl :

1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_ba3.\*  
4: gb\_in1.\*  
5: gb\_in2.\*  
6: gb\_in3.\*  
7: gb\_om.\*  
8: gb\_ov.\*  
9: gb\_pat1.\*  
10: gb\_pat2.\*  
11: gb\_ph.\*  
12: gb\_pl1.\*  
13: gb\_pl2.\*  
14: gb\_pl3.\*  
15: gb\_pl4.\*  
16: em\_ba1.\*  
17: em\_ba2.\*  
18: em\_fun.\*  
19: em\_htgo\_hum.\*  
20: em\_htgo\_inv.\*  
21: em\_htgo\_rod.\*  
22: em\_htg\_hum1.\*  
23: em\_htg\_hum2.\*  
24: em\_htg\_hum3.\*  
25: em\_htg\_hum4.\*  
26: em\_htg\_hum5.\*  
27: em\_htg\_hum6.\*  
28: em\_htg\_hum7.\*  
29: em\_htg\_hum8.\*  
30: em\_htg\_inv1.\*  
31: em\_htg\_inv2.\*  
32: em\_htg\_other.\*  
33: em\_htg\_rod.\*  
34: em\_hum1.\*  
35: em\_hum2.\*  
36: em\_hum3.\*  
37: em\_hum4.\*  
38: em\_hum5.\*  
39: em\_hum6.\*  
40: em\_hum7.\*  
41: em\_in.\*  
42: em\_om.\*  
43: em\_or.\*

44: em\_ov.\*  
45: em\_pat.\*  
46: em\_ph.\*  
47: em\_pl.\*  
48: em\_rod.\*  
49: em\_sts.\*  
50: em\_sy.\*  
51: em\_un.\*  
52: em\_vi.\*  
53: gb\_sts1.\*  
54: gb\_sts2.\*  
55: gb\_sts3.\*  
56: gb\_sy.\*  
57: gb\_un.\*  
58: gb\_vil.\*  
59: gb\_vil2.\*  
60: gb\_htg1.\*  
61: gb\_htg2.\*  
62: gb\_htg3.\*  
63: gb\_htg4.\*  
64: gb\_htg5.\*  
65: gb\_htg6.\*  
66: gb\_htg7.\*  
67: gb\_htg8.\*  
68: gb\_htg9.\*  
69: gb\_htg10.\*  
70: gb\_htg11.\*  
71: gb\_htg12.\*  
72: gb\_htg13.\*  
73: gb\_htg14.\*  
74: gb\_htg15.\*  
75: gb\_htg16.\*  
76: gb\_htg17.\*  
77: gb\_htg18.\*  
78: gb\_htg19.\*  
79: gb\_htg20.\*  
80: gb\_htg21.\*  
81: gb\_htg22.\*  
82: gb\_htg23.\*  
83: gb\_htg24.\*  
84: gb\_htg25.\*  
85: gb\_pr1.\*  
86: gb\_pr2.\*  
87: gb\_pr3.\*  
88: gb\_pr4.\*  
89: gb\_pr5.\*  
90: gb\_pr6.\*  
91: gb\_pr7.\*  
92: gb\_pr8.\*  
93: gb\_pr9.\*  
94: gb\_rod.\*  
95: gb\_rod2.\*  
96: gb\_in4.\*  
97: gb\_pr10.\*  
98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	25	100.0	128	2	ECOBIOB
2	25	100.0	1041	9	AR029499
3	25	100.0	1041	9	AR034916
4	25	100.0	1084	9	A11530
5	25	100.0	5526	2	AF250776
6	25	100.0	5793	2	ECOBIO
7	25	100.0	5872	9	A38246
8	25	100.0	5872	9	A38251
					M27731 Escherichia
					AR029499 Sequence
					AR034916 Sequence
					A11530 BioB gene o
					AF250776 Unculture
					J04423 E.coli 7,8-
					A38246 Sequence 1
					A38251 Sequence 6

```

9 25 100.0 5872 9 A93674
10 25 100.0 5872 9 A93679
11 25 100.0 5872 9 AR101809
12 25 100.0 5872 9 AR101810
13 25 100.0 11022 1 AE000180
14 25 100.0 13501 1 AE005258
15 25 100.0 297816 2 AF002553
16 24 96.0 1121 10 E00893
17 20.8 83.2 12891 1 AE004192
18 20.2 80.8 1041 3 EBIOTOP2
19 20.2 80.8 7215 3 SHABIO
20 19.8 79.2 8227 2 AF248314
21 19.8 79.2 185738 76 AC084015
22 19.8 79.2 218375 67 AC022397
23 19.8 79.2 302558 72 AC063928
24 19.8 79.2 303717 72 AC063930
25 19.2 76.8 22329 68 AC014347
26 19.2 76.8 38000 68 AF025442
27 19.2 76.8 110000 84 HSMX1_1
28 19.2 76.8 138938 61 AC010111
29 19.2 76.8 176928 88 AC084239
30 19.2 76.8 181636 4 AC008224
31 19.2 76.8 184304 62 AC011822
32 19.2 76.8 30050 93 HSMX1A
33 19.2 76.8 340000 92 HS21C084
34 18.8 75.2 550 94 AF158246
35 18.8 75.2 107024 66 AC021088
36 18.6 74.4 1029 53 CNS06NEL
37 18.6 74.4 4037 13 ATH249204
38 18.6 74.4 122877 88 AC034199
39 18.4 73.6 168091 64 AC016895
40 18.4 73.6 190000 89 AF213884S1
41 18.4 73.6 192368 66 AC021120
42 18.2 72.8 2018 4 AE002676
43 18.2 72.8 2021 65 AC019735
44 18.2 72.8 2094 8 XLRNAHE
45 18.2 72.8 3429 94 AB029919

```

## ALIGNMENTS

```

RESULT 1
ECOBIOB ECOBIOB 128 bp DNA BCT
LOCUS Escherichia coli biotin (bioB) gene, early terminator region.
DEFINITION M2731
ACCESSION M2731
VERSION M2731.1 GI:341755
KEYWORDS bioB gene; biotin.
SOURCE Escherichia coli (strain K-12) DNA.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 128)
AUTHORS Nath,S.K.
TITLE Attenuation of transcription of biotin genes in Escherichia coli
JOURNAL Can. J. Microbiol. 34 (12), 1288-1296 (1988)
MEDLINE 89167942
FEATURES
    source
        1..128
            /organism="Escherichia coli"
            /strain="K-12"
            /db_xref="taxon:562"
        4..128
            /gene="bioB"
        4..9
            /gene="bioB"
        20..>128
            /gene="bioB"
        103..121
            /gene="bioB"
        /note="early terminator"
BASE COUNT 31 a 28 c 35 t

```

```

ORIGIN
Query Match 100.0%; Score 25; DB 2; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctggatctgctgtttgaagcg 25
|||||
Db 18 GTTGCTGGATCTGCTGTTTGAAGCG 42

RESULT 2
AR029499
LOCUS AR029499 1041 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
    SOURCE
        Unknown.
    ORGANISM
        Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
    Location/Qualifiers
        source
            1..1041
                /organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctggatctgctgtttgaagcg 25
|||||
Db 57 GTTGCTGGATCTGCTGTTTGAAGCG 81

RESULT 3
AR034916
LOCUS AR034916 1041 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
    SOURCE
        Unknown.
    ORGANISM
        Unclassified.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
    Location/Qualifiers
        source
            1..1041
                /organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctggatctgctgtttgaagcg 25
|||||
Db 57 GTTGCTGGATCTGCTGTTTGAAGCG 81

RESULT 4
QY 1 gttgctggatctgctgtttgaagcg 25
|||||
Db 57 GTTGCTGGATCTGCTGTTTGAAGCG 81

```

All1530  
 LOCUS All1530 1084 bp DNA PAT 09-FEB-1994  
 DEFINITION BioB gene of E.coli with primers.  
 ACCESSION All1530  
 VERSION All1530.1 GI:490218  
 KEYWORDS  
 SOURCE Escherichia coli.  
 ORGANISM Escherichia coli.  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Escherichia.  
 1 (bases 1 to 1084)  
 Patent: GB 2216530-A 16 11-OCT-1989;  
 Location/Qualifiers  
 1..1084  
 /organism="Escherichia coli"  
 /db\_xref="taxon:562"  
 24..1064  
 /gene="BioB"  
 24..1064  
 /gene="BioB"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA00965.1"  
 /db\_xref="GI:490219"  
 /db\_xref="SWISS-PROT:PI2996"  
 /translation="MAHRPRTLQVTELFKEKPLDLLLFEAQVHROHFDPRQVQVST  
 LLSIKTGACPEDCKYCPOSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW  
 KNPNERDMPYLEQWVGVDLGLACMTLGLTLESQAORLANAGLDYNNHNLDTSPF  
 YGNITRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLIQLANLPTPEF  
 VPINMLVKVGTPLADNDVDADFIRTIARIMPTSVYRLSAGREOMNEQOAMC  
 FMAGNSIFYGCKLLTTPNEEDKDLQFRKGLGNPQQTAVLAGDNEQQRLQALMT  
 PDTDEYNAAL"  
 271 a 286 c 318 g 209 t  
 BASE COUNT  
 ORIGIN  
 Query Match 100.0%; Score 25; DB 9; Length 1084;  
 Best Local Similarity 100.0%; Pred. No. 0.27;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 gttgctgagctgctgtttgaagcg 25  
 ||||||||||||||||||||  
 Db 80 GTTGTGGATCTGCTTTGAAGCG 104  
 RESULT 5  
 AF250776  
 LOCUS AF250776 5526 bp DNA BCT 31-JAN-2001  
 DEFINITION Uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in  
 modC-bioA intergenic region, DAPA-aminotransferase BioA (bioA),  
 biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and  
 biotin biosynthesis protein BioC (bioC) genes, complete cds; and  
 dethiobiotin synthetase BioD (bioD) gene, partial cds.  
 AF250776  
 AF250776.1 GI:12620124  
 uncultured bacterium pCosHE2.  
 uncultured bacterium pCosHE2.  
 Bacteria; environmental samples.  
 1 (bases 1 to 5526)  
 Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streit, W.R.  
 Direct cloning from enrichment cultures, a reliable strategy for  
 isolation of complete operons and genes from microbial consortia  
 Appl. Environ. Microbiol. 67 (1), 89-99 (2001)  
 20575196  
 1113432  
 2 (bases 1 to 5526)  
 Entcheva, P., Liebl, W. and Streit, W.R.  
 Direct Submission  
 Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet  
 Goettingen, Grisebachstr. 8, Goettingen 37077, Germany  
 Location/Qualifiers  
 1..5526  
 /organism="uncultured bacterium pCosHE2"  
 /db\_xref="taxon:143797"  
 /clone="pCosHE2"  
 /notes="unknown organism, cosmid clone derived from  
 environmental consortium"  
 complement(52..528)  
 /note="ORF1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical 17.1 kDa protein in modC-bioA  
 intergenic region"  
 /protein\_id="AAC60577.1"  
 /db\_xref="GI:12620125"  
 /translation="MKLISNDLRDQSKLPHRVFNGMGYDGNISPHLAWDDVPAGTK  
 SFVTCYDPDAPTQSGWHVYVNLPAFTRVLPGQFSGVLVAMPDGVLTQTRDTPGKAG  
 YDGAAPPKGETHRYIFTVHALDIERIDVDGAGAMVGFNVHFSLASASITAFMS"  
 complement(587..1876)  
 /gene="bioA"  
 complement(587..1876)  
 /gene="bioA"  
 /notes="7,8-diaminopelargonic acid  
 synthetase-aminotransferase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="DAPA-aminotransferase BioA"  
 /protein\_id="AAC60578.1"  
 /db\_xref="GI:12620126"  
 /translation="MTTDDLAFDORHIHPYTSMTSPLPVYPVSAEGCELILSDGRR  
 LVDGSSWAAITNGYNHNPOLNAFMSQIDAMSHVMFGGITHAPATELCRKLVMTPOP  
 LKCVFLADGSAVAVEAMKALQYNQAKGEQRELTFRNGYHGDTFGMSVCDPDNS  
 MSLWKGYPENLFAFAPQSRMDGDEWDERMVGFAKMAAHEHIAAVTIEPIVVGAG  
 GMRWHPKWLKRIKICDREGILITADEIATFGRTGKLFACFAEIAPIADILCLGKAL  
 TGGTMTLSATLTTRVEAETISNGEACFMHGPTFMGNPLACAAASALILESGMQQ  
 QVADIEVQLREQLAPARDAEMVADVRLGAIGVETTHPVNMAALQKFVEQGVWIRP  
 FGKLIYMPYIILPQQLRLTAAVNRAVDQETFFCQ"  
 1963..3003  
 /gene="bioB"  
 1963..3003  
 /gene="bioB"  
 /codon\_start=1  
 /transl\_table=11  
 /product="biotin synthase BioB"  
 /protein\_id="AAC60579.1"  
 /db\_xref="GI:12620127"  
 /translation="MAHRPRTLQVTELFKEKPLDLLLFEAQVHROHFDPRQVQVST  
 LLSIKTGACPEDCKYCPOSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW  
 KNPNERDMPYLEQWVGVDLGLACMTLGLTLESQAORLANAGLDYNNHNLDTSPF  
 YGNITRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLIQLANLPTPEF  
 VPINMLVKVGTPLADNDVDADFIRTIARIMPTSVYRLSAGREOMNEQOAMC  
 FMAGNSIFYGCKLLTTPNEEDKDLQFRKGLGNPQQTAVLAGDNEQQRLQALMT  
 PDTDEYNAAL"  
 3000..4154  
 /gene="bioF"  
 3000..4154  
 /gene="bioF"  
 /notes="8-amino-7-oxononanoate synthase"  
 /codon\_start=1  
 /transl\_table=11  
 /product="KAPA synthetase BioF"  
 /protein\_id="AAC60580.1"  
 /db\_xref="GI:12620128"  
 /translation="MSWQEKINAALDARRAADLRRYPVPAQAGRWLVADDRQYLN  
 SNDYILGLSHHPQIIIRAWQGAQFGISGSHVSGYSVHQAELAEELAEWLGYSRA  
 LLFTSGFAANQAVIIRAMMAKEDRTADRLSHASLLLEAASLSPSQLRRAHNDVTHLAR  
 LLASPCPGQOOLVVTGCVFSGMDGDSAPLAEICQVTOHNGWLMVDHAGHTVTEGQGR  
 SCWLOKVPPELLVITFGKFGVSGAIVLCSTVADYLLQFAEHLIYSTMPAQAL  
 RASLAVIRDEGDARREKLVSLIAHFRAGVQDLPTPLADSCSAIOLIVGNSRALQL  
 AEKLRQOGCWTAIRPPTVPAGTARLRLTLTAHEMQDIDRLLEVLHGNG"  
 4141..4896  
 /gene="bioC"  
 4141..4896  
 /gene="bioC"  
 /transl\_table=11  
 /product="KAPA synthetase BioF"  
 /protein\_id="AAC60580.1"  
 /db\_xref="GI:12620128"  
 /translation="MSWQEKINAALDARRAADLRRYPVPAQAGRWLVADDRQYLN  
 SNDYILGLSHHPQIIIRAWQGAQFGISGSHVSGYSVHQAELAEELAEWLGYSRA  
 LLFTSGFAANQAVIIRAMMAKEDRTADRLSHASLLLEAASLSPSQLRRAHNDVTHLAR  
 LLASPCPGQOOLVVTGCVFSGMDGDSAPLAEICQVTOHNGWLMVDHAGHTVTEGQGR  
 SCWLOKVPPELLVITFGKFGVSGAIVLCSTVADYLLQFAEHLIYSTMPAQAL  
 RASLAVIRDEGDARREKLVSLIAHFRAGVQDLPTPLADSCSAIOLIVGNSRALQL  
 AEKLRQOGCWTAIRPPTVPAGTARLRLTLTAHEMQDIDRLLEVLHGNG"  
 4141..4896  
 /gene="bioC"  
 4141..4896  
 /gene="bioC"

```

/note="reaction step prior to synthesis of pimeloyl-CoA"
/codon_start=1
/transl_table=11
/product="biotin biosynthesis protein BioC"
/protein_id="AAG60581.1"
/db_xref="GI:12620129"
/translation="MATVNRKQAAIAAFGRAAALYEQHADLQROASADVLAMLPKRYKT
HVLDAAGCGQMTWRHWRHAQVYALDLSPPMLVQARQDAADHYLAGDIESIPLATA
TFDLANSNAVQWGNLSALRELYRVRSKGVAFYTLVQGLPELQAWQAVDERP
HANRFLPDEIQSLNGVYHQHIQPIITLWFDLALSAMRSLSKIGATHLHEGRDPRIL
TRSQLRLQLANFQQQGRYPLTYHLFLGVARE"
4883..>5526
/gene="bioD"
CDS
4883..>5526
/gene="bioD"
/note="DTB synthetase"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase BioD"
/protein_id="AAG60582.1"
/db_xref="GI:12620130"
/translation="MLVSKRYFVTGDTVEGKTVASCALLQAAKAAGYRTAGYKPVAS
GSEKTPGLNSDALQORNSLQLDYATVNPYTEAPTSPIISAQEGRPTESSVMS
SGRALBQADWLVESAGGWFTPLSDTFEADWVTOEQLPVILVGVKLGICINHAML
TAQAIQHAGITLAGWANDVTPPGKRRHAEYITTLTRMIPAPLLGEIPWLAENPE"
BASE COUNT 1274 a 1507 c 1567 g 1178 t
ORIGIN

Query Match 100.0%; Score 25; DB 2; Length 5526;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 qtgtcgatctgctgttgaagcg 25
|||||
Db 2019 GTTGTGATCTGCTGTTGAAGCG 2043

RESULT 6
ECOBIO 5793 bp DNA BCT 28-FEB-1994
LOCUS E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
DEFINITION (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC
protein, and dethiobiotin synthetase (bioD), complete cds.
ACCESSION J04423
VERSION J04423.1 GI:145422
KEYWORDS 7,8-diamino-pelargonic acid aminotransferase;
7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
bioC gene; bioD gene; bioF gene; biotin synthetase; dethiobiotin
synthetase.
SOURCE Escherichia coli (strain K-12) DNA.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5793)
AUTHORS Otsuka,A.J., Buoncristiani,M.R., Howard,P.K., Flamm,J. and
Johnson,O.
TITLE The Escherichia coli biotin biosynthetic enzyme sequences
predicted
JOURNAL J. Biol. Chem. 263, 19577-19585 (1988)
MEDLINE 89066784
COMMENT Draft entry and computer-readable sequence [1] kindly submitted by
A.Otsuka, 09-NOV-1988.
FEATURES
source
Location/Qualifiers
1..5793
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
complement(98..574)
/note="ORF 1"
/codon_start=1
/transl_table=11
/protein_id="AAA23513.1"

```

```

/db_xref="GI:455168"
/translation="MKLISNDLRDGLKPLRRHVFNGMGYDGDNTSPHLAMDVDPAGTK
SFVTCYDPAFTSGTGMWVWVVPADTRVLPOCFGSLVAMPDGLQTRDFGKTG
YDGAAPPKGETHRITFTVHALDIERIDVDEGAAMVGFNVHFHSLASATAMFS"
complement(633..1925)
/gene="bioA"
CDS
complement(633..1925)
/gene="bioA"
/codon_start=1
/transl_table=11
/product="7,8-diamino-pelargonic acid aminotransferase"
/protein_id="AAA23514.1"
/db_xref="GI:457106"
/translation="MTTDDLAFDQQRHIWHPYTSMTSPLVPYVVSAGCECILSDGRR
LVDGSSWMAAIIHGYNHNPQLNAAMKSOIDAMSHVMFGGITHAPAIEICRLVAMSGRN
ALECVFLADSGSVAVEAVKALOVYQAKGEARORFLTFRNGYHGDYFGAMSVCDPDN
SMHSLWKGYLPENLPAFAPAPQSRMDGEMDERMDVGFARLMAAHRHEIAAVITEPIVOGA
GOMRYHPWLKRIKICDREGILLIADIEATIGPRTGKLPFACBHAETAPDILCLGKA
LTGQMTLSATLITREVAETISNGEAGCFMHGPTFMGNPLACAAASLASLILIESGDMQ
QVADIYOLREQLAPARDAEMVADRVILGIVGVETHPVYMAALQKFFVEQGVWIR
PFGKILYLPPIYIILPOOLRLTAAVNRAVQDETFFCQ"
2012..3052
/gene="bioB"
CDS
2012..3052
/gene="bioB"
/codon_start=1
/transl_table=11
/product="biotin synthetase"
/protein_id="AAA23515.1"
/db_xref="GI:145425"
/translation="MAHRPRWTLNQVTELFKEKPLDLLLFEAQOVHROHEDPROQVST
LLSITGACPEDCKYCPTSRITKGLAEERLMEVEQVLESARKAKAGSTRFCMGAAW
KNHTERDMPYLEQWGVQKAMGLACMTGLTSLSEQARQANGLDYNHNLDTSPF
YGNITRTYQERLDLEKVRDAGIKVCSGIVGLGETVKDRAGLLQLLNLPTPES
VPINLVKVKGTPLADNDVDADFIRTIIVARIMMPTSYVRLSAGREMQNEQAMC
FMAGANSIFYGCKLLITTPNEEDKDLQFLKRLGILNPQQTAVLAGDNEQQRLEQALMT
POTDEYNAAL"
3049..4203
/gene="bioF"
CDS
3049..4203
/gene="bioF"
/codon_start=1
/transl_table=11
/product="7-keto-8-amino pelargonic acid synthetase"
/protein_id="AAA23516.1"
/db_xref="GI:145426"
/translation="MSWQEKINAAIDARRAADALRRYPVPAQAGRWLVADDDRYLNF
SSNDYLGISHHPQIIIRAWQOQEQFGIGSGSGHVSYSVYHQALEELAEWLGYSRA
LILFISGFAANQAVIAAMAKEDRIAADRLSHASLLEAASLSPQLRRFAHNDVTHLAR
LLASPCQGVVTVTEGVFSMDGDSAPLAEIOQVTOHQHGLMVDADHGTGIVGQGRG
SCHLOKVPPELLVYTFGKFGVSGNAVLCSTVADYLLQFAHLLIYSTMPPAQOAL
RASLAVIRSDGSDARREKLAALITFRAGVQDLPTLADSCSAIQPLIVGNSRALQL
AEKLRQCGCWTAIRPPTVPAGTARLRLTLTAHEMQDIDRLLEVLHGNG"
4190..4945
/gene="bioC"
CDS
4190..4945
/gene="bioC"
/codon_start=1
/transl_table=11
/protein_id="AAA23517.1"
/db_xref="GI:145427"
/translation="MATVNRKQAAIAAFGRAAALYEQHADLQROASADVLAMLPKRYKT
HVLDAAGCGQMTWRHWRHAQVYALDLSPPMLVQARQDAADHYLAGDIESIPLATA
TFDLANSNAVQWGNLSALRELYRVVPKGVAFYTLVQGLPELQAWQAVDERP
HANRFLPDEIQSLNGVYHQHIQPIITLWFDLALSAMRSLSKIGATHLHEGRDPRIL
TRSQLRLQLANFQQQGRYPLTYHLFLGVARE"
4938..5597
/gene="bioD"
CDS
4938..5597
/gene="bioD"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase"

```



/protein\_id="AAA23518.1"  
 /db\_xref="GI:145428"  
 /translation="MSKRYFVTGTDTEVGKTVASCALLOAKRAGRYTAGYKPVASGS  
 EKTPLGRNSDALALORSSLDYATVNPYTFAPTSPIHISAQGRPIESLSVMSAG  
 LRALEQADWLVVEGAGWFTPLSDTFTFADWVTEQLPVILVGVKLGICINHAMLTA  
 QVIOHAGLTLAGVANDVTPPKRRAEYMTLTRMIPRCWERSPGIKQTKRQPPES  
 T"  
 BASE COUNT 1363 a 1554 c 1631 g 1245 t  
 ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match 100.0%; Score 25; DB 2; Length 5793;  
 Best Local Similarity 100.0%; Pred. No. 0.31;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgttgatctgtctgtttgaagcg 25  
 |||TTTTTTTTTTTTTTTTTTTT|  
 Db 2068 GTTGTGATCTGCTGTTGAAGCG 2092

RESULT 7  
 A38246  
 LOCUS A38246 5872 bp DNA PAT 05-MAR-1997  
 DEFINITION Sequence 1 from Patent WO9408023.  
 A38246  
 VERSION A38246.1 GI:2294844  
 KEYWORDS  
 SOURCE Escherichia coli.  
 ORGANISM Escherichia coli.  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Escherichia.

REFERENCE 1 (bases 1 to 5872)  
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.  
 TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN  
 JOURNAL Patent: WO 9408023-A 1 14-APR-1994;  
 LONZA AG (CH)

COMMENT  
 Other publication PL 308301 950724  
 Other publication CA 2145400 940414  
 Other publication AU 4820293 940426  
 Other publication HU 71781 960228  
 Other publication SK 42095 951108  
 Other publication CZ 9500809 950913  
 Other publication FI 951547 950331  
 Other publication JP 8501694T 960227.  
 Location/Qualifiers

FEATURES  
 source  
 1..5872  
 /organism="Escherichia coli"  
 /strain="DSM498"  
 /db\_xref="taxon:562"  
 /clone="PBO30A-15/9"  
 1..96  
 /function="PROMOTER PTAC"  
 /evidence=experimental  
 23..28  
 45..49  
 /standard\_name="PROMOTER PTAC"  
 /evidence=experimental  
 105..119  
 /standard\_name="BIOB RBS NO.9"  
 /evidence=experimental  
 117..1157  
 /gene="BIOB"  
 117..1157  
 /gene="BIOB"  
 /codon\_start=1  
 /transl\_table=11  
 /number=1  
 /evidence=experimental  
 /product="BIOTIN SYNTHASE"  
 /protein\_id="CAA02325.1"  
 /db\_xref="GI:2294845"  
 /translation="MAHRPRWTLISQVTELFKPLLDLLEFAQQVHRQHFDPQVQVST  
 LLSIKTGACPEDCKYCPQSSRYKTYGLEARLMEVQVLESARKAKAAGSTRFCMGAAW

promoter

-35\_signal  
 -10\_signal

RBS

gene

CDS

stem\_loop  
 terminator

BASE COUNT 1318 a 1552 c 1695 g 1307 t  
 ORIGIN

Query Match

100.0%; Score 25; DB 9; Length 5872;

KNPHERMPYLEQMVQGVKAMGLEACMTLGLTISEQAQRIANAGLDYYNNHLDTSPEF  
 YGNIITRTYQERLDTLEKVRDAGIKVCGSGIVGETVKDRAGILLQLANLPTPPES  
 VFINMLVKVKGTPPLADNDVDADFRTIATVARIIMPTSYVRLSACREQNNQTOAMC  
 FMAGINSIYGVCKLLTTPPEEDKDLQFLFKCLNPOQTAVLAGDNEQQORLEQALMT  
 PDPDEYNAAL  
 2284..2297  
 /standard\_name="BIOC RBS"  
 2295..3050  
 /gene="BIOC"  
 2295..3050  
 /gene="BIOC"  
 /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"  
 /codon\_start=1  
 /transl\_table=11  
 /number=3  
 /product="PROTEIN"  
 /protein\_id="CAA02326.1"  
 /db\_xref="GI:2294846"  
 /translation="MATVANKQATAAFGRAAHYEQHADLQROSDADALLAMLPORKYT  
 HVLDAAGCGPMWSRHRERHAQVATLDLSPMLVQARQKDAADHYIAGDIESPLATA  
 TFDLAWNSLAVQWCGNLSTALRELYRVVRPKGVVATTLVQCSLPLHQAWQAVDERP  
 HANRFLPDEIEQSLNGVHYQHIIQPIITLWFDLALSAMRSLAGIGATHLHEGRDPRIL  
 TRSOLRLOLAMPQQQGRYPLTYHLFLGVARE"  
 3742..3752  
 /standard\_name="BIOA RBS"  
 3750..5039  
 /gene="BIOA"  
 3750..5039  
 /gene="BIOA"  
 /standard\_name="S-ADENOSYL-L-METHIONINE:  
 8-AMINO-7-OXONONANOATE AMINOTRANSF."  
 /EC\_number="2.6.1.62"  
 /codon\_start=1  
 /transl\_table=11  
 /number=5  
 /evidence=experimental  
 /product="DAPA SYNTHASE"  
 /protein\_id="CAA02327.1"  
 /db\_xref="GI:2294847"  
 /translation="MTTDLAFDQRHIWHPYTSPLVPVVSAGCELILSDGRR  
 LVDGSGYMWAAIHGYNHNPOLNAAMKSOIDAMSHVMFGGITTHAPATELCRKLVAWTPQ  
 LECVFLADSGSVAVEVAMKMALOYWQAKGEARQRLTPRNGYHGDTFGMSVCDPDNS  
 MSLWKGYLPENLPAPQSRMDGDEWDMQFARLMAAHRHETAAVIEPIVQVAG  
 GMRMHPWKIRKICDREGILLIADETATGFGRTKGLFACEHAETAPDILCLGKAL  
 TGGTMTLSATLTTRVAETISNGEACGFMHGPTMGNPLACAAANASLAILESSDQOQ  
 QVADIEVOLREQLAPARDAEMVADVRLGACIGVETTHPVNMAALQKFEVQGVWIRP  
 FGKLIYLMPPYIILPQQLRQLTAAVNRAVQDETFCQ"

5088..5100  
 /standard\_name="ORFI RBS"  
 5098..5574  
 /gene="ORFI"  
 5098..5574  
 /gene="ORFI"  
 /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"  
 /codon\_start=1  
 /transl\_table=11  
 /number=6  
 /evidence=experimental  
 /product="PROTEIN"  
 /protein\_id="CAA02328.1"  
 /db\_xref="GI:2294848"  
 /translation="MKLISNLDLRGDKLPHRHFVNGMGYDGNISPHLAWDDVPAGTK  
 SFVVTCTDPDTPATGSGMWHVNVNLPADTVLPQGGSLVAMPDGVLTQRTDFGKTG  
 YDGAAPKGETHRYIFTVHALDIERIDVDEGASGMVGVNFVHFHSLASATAMFS"  
 5583..5605  
 5583..5644  
 /standard\_name="RHO-INDEPENDENT TRANSCRIPTIONAL  
 TERMINATOR"

Best Local Similarity 100.0%; Pred. No. 0.31; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgtggtatctgctgtttgaagcg 25  
|||||

Db 173 GTTGTGGATCTGCTTTGAAGCG 197

## RESULT 8

A38251 A38251 5872 bp DNA PAT 05-MAR-1997  
DEFINITION Sequence 6 from Patent WO9408023.  
ACCESSION A38251  
VERSION A38251.1 GI:2294849

KEYWORDS Escherichia coli.  
SOURCE Escherichia coli  
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

REFERENCE 1 (bases 1 to 5872)  
AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.  
TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN  
JOURNAL Patent: WO 9408023-A 6 14-APR-1994;

COMMENT LONZA AG (CH)  
Other publication PL 308301 950724  
Other publication CA 2145400 940414  
Other publication AU 4820293 940426  
Other publication HU 71781 960228  
Other publication SK 42095 951108  
Other publication CZ 9500809 950913  
Other publication FI 951547 950331  
Other publication JP 8501694T 960227.

## FEATURES

source  
1..5872  
/organism="Escherichia coli"  
/strain="DSM498"  
/db\_xref="taxon:562"  
/clone="PBO30A15-9"  
1141..1156  
/standard\_name="BIOF RBS"  
1154..2308  
/gene="BIOF"  
1154..2308  
/gene="BIOF"  
/EC\_number="2.3.1.47"  
/codon\_start=1  
/transl\_table=11  
/number=2  
/evidence=experimental  
/product="KAPA SYNTHASE"  
/protein\_id="CAA02329.1"  
/db\_xref="GI:2294850"  
/translation="MSWQEKI NAALDARRADALRRRYPVAQAGRWLVADRDQYLN  
SSNDYLGISHUPQIIRAWQGAOFGIGSGSHVSGYVVVHQALEELAEWLGY  
LLFISFAANQAVIAAMAKEDRIAADRLSHASLEAASLPSOLRRFAHNDVTHAR  
LLASPCQQWVVEGVESMDGDSNPLAEIQVTCQINGWLVDDAHTGTVIGEGRG  
SCWQKVPPELLVTFGKGVSGAALVCSSTVADILQFAHLLIYSMPAQAQAL  
RASLAVIRSDGDAAREKLAALITFRAGVODLPFTLADSCSAIOPLIVGDSRALQ  
AEKLRQCGWVTAIRPPTVPAGTARLRLTLTAHEMODIDRLLEVLHGNG"  
3030..3045  
/standard\_name="BIOD RBS"  
3043..3753  
/gene="BIOD"  
3043..3753  
/standard\_name="BIOD"  
/EC\_number="6.3.3.3"  
/codon\_start=1  
/transl\_table=11  
/number=4  
/evidence=experimental  
/product="DTB SYNTHASE"

## RBS

gene

## CDS

/protein\_id="CAA02330.1"  
/db\_xref="GI:2294851"  
/translation="MSKRYFVTGDTTEVKTIVASCALLOAKAAGYRTAGYKPVASGS  
EXTPEGLRNSDALALQRNSSLQLDYATVNPYTFRAEPTSPHILISAGEGRPIESLYMSAG  
LRALEQQADWLVLEGAGWFTPLSDTFTFDWVTOEQLPVLIVGVKLGCIINHAMLTA  
QVIOHAGLTLAGWAVANDVTPPKRHAETMTLTRMIPAPLIGEIPWLAENPENNAATGK  
YINLAFVDASTLGTFTSRL"

BASE COUNT 1318 a 1552 c 1695 g 1307 t  
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;  
Best Local Similarity 100.0%; Pred. No. 0.31;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgtggtatctgctgtttgaagcg 25  
|||||

Db 173 GTTGTGGATCTGCTTTGAAGCG 197

## RESULT 9

A93674 A93674 5872 bp DNA PAT 22-JAN-2000  
DEFINITION Sequence 1 from Patent EP0798384.  
ACCESSION A93674  
VERSION A93674.1 GI:6741862

KEYWORDS Escherichia coli.  
SOURCE Escherichia coli  
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Escherichia.

REFERENCE 1 (bases 1 to 5872)  
AUTHORS Birch, O. and Brass, J.  
TITLE Biotechnological method of producing biotin  
JOURNAL Patent: EP 0798384-A 1 01-OCT-1997;  
LONZA AG (CH)

## FEATURES

source  
1..5872  
/organism="Escherichia coli"  
/strain="DSM498"  
/db\_xref="taxon:562"  
/clone="PBO30A-15/9"  
1..96  
/function="PROMOTER PTAC"  
/evidence=experimental  
23..28  
/standard\_name="PROMOTER PTAC"  
45..49  
/evidence=experimental  
105..119  
/standard\_name="BIOB RBS NO.9"  
/evidence=experimental  
117..1157  
/gene="BIOB"  
117..1157  
/gene="BIOB"  
/codon\_start=1  
/transl\_table=11  
/evidence=experimental  
/product="BIOTIN SYNTHASE"  
/protein\_id="CAB69590.1"  
/db\_xref="GI:6741863"  
/translation="MAHRPRWTLTSQVTELFEXPLDLLFEAQVHRQHFDPQVQVST  
LLSITKTCACPEDCKYCPSSRYKTGLEAERLMEVLESAKKAAGSTRFCMGAAW  
KNPHERDMPYLEOMVQGVKAMGLEACMTGLTSESQAORLANAGLDYVNHLDTSPEF  
YGNLITTRYQERLDTLEKVRDAGIKVCSGGIVGLETYKDRAGLLLOLANLPTPES  
VPINLITRYKVTPLADNDVDADFRTIATVAKIMMPTSYVRLSAGREOMNEQTQAMC  
FMAGANSIFYGCKLLTTPNEPDKQLQLFKRLGRLNPQQTAVLAGDNEQQORLEQALMT  
PDTDEYNNAAAL"  
2284..2297  
/standard\_name="BIOC RBS"  
2295..3050  
/gene="BIOC"

```

CDS
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVKNQAIAAAGRAAHYEQHADLQKQADALLAMLPORKYT
HYLDAGCGPMWSHRERHAQVATLDDSPMLVOAKQKDAADHYLAGDIESLPLATA
TFDLAWSNLAVOWCGNISTALRELYRVPRKGVVAFVTLVQGSUPELHQAWQVDERP
HANRFLPDETEQSLNGVYQHIIQPIITLFWDDALSAMRSLKGTIGATHLHEGRDPRIL
TRSQLRLQLAWPQQQGRYPLTYHLFLGVARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONANOATE AMINOTRANSF."
/EC_number="2.3.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDDLAFDQRIWHPYTSMTPSPVYVVSABGCELIISDGR
LVDGMSWMAIHHYNNPQLNAAMKQSDIMASHVMFGGITHAPEIELCKRLVAMTPO
LECVELADSGSVAVEAMKALVOYAKGEARQFLFRNGYHGDTPGAMSCDPPONS
MHSWKGYLPENLPAPAPQSMQDGEWDERMVGFARLMAARHEIAVILEPIVQAG
GRMYHPENLKRKICDREGILLIADIATGRTKLFACAEIAPDILCIGKAL
TGGMTLSATLTREVAETISNGEAGCFMHPGTFMGNPLACAAANALLESQDWQO
QVADIEVQLREQLAPADAEMADVRLVIGAVYVETHPYNNMAALQKFFVEQGVWIRP
FGKLIYLPYPIIILPOQLRLTAANRAVODETFQC"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDGLPDRHVPNGYDGDNIISPHLANDDDVPAGTK
SFVYTCIDPDPTGSGWVWVNNLPADTRVLPQFGSGGLVAMPDGVLTQRTDFGKITG
YDGAAPRGETHRYHIFTVHALDIERIDVDEGASGVGVNHFHSLASITAMFS"
5583..5605
/stem_loop
terminator
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctgactctgtttgaagc 25
|||||
Db 173 GTTGCTGGATCTGCTGTTGAAGCG 197

RESULT 10
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS Sequence 6 from Patent EP0798384.

ACCESSION A93679 GI:6741867
VERSION A93679.1
KEYWORDS Escherichia coli.
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O. and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
source Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQEKINAALDARRADARRRRYPVQAQGRWLVDROQYLN
SSNDYLGSLHHFQIIRAWQGAPEFGISGGSGHVSQSVVHQALAEELAEWLGYSA
LIFISGAANQAVIAAMAKEDRIAADRLSHASLLEAASLSPQLRRFAHNDVTHLAR
LIASPCGQOMVVTGCVFSGMDGSAFLAEIQVTOHNGMLWMDAAGTGVIGEOGRG
SCWLQVKKPELLVVTGKFGYSGAVALCVSSVADYLLQFAHLLIYSTSMPPAQAL
RASLAVIRDEGADARREKLAALITFRAGVQDLPTLADSCSAIQPLIVGDSRALQL
AEKLRQGCWTAIRPPVPAFTARLRLTLTAHEMQDIDRLLEVLHGNG"
3030..3045
/standard_name="BIOD RBS"
3043..3753
/gene="BIOD"
3043..3753
/gene="BIOD"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTF SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYFVTGTDTEVGTVCALLQAAKAAAGVTAQYKPVASGS
EKTPEGLRNSDALQORNSLQLDVATVNPYTAFTSPHIIAQEGRPESLVMSAG
LRALQEQADWLVVEGAGWFTPLSTFTFADWVTOQLPVLVVGKLGCCINHAMLTA
QVLIQAGLTLAGVANDVTPGKRAEYNTLTTRMIPAPLLGEIPLWLAENPENAAATGK
YINLAFVDASTLGTFTSRL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctgactctgtttgaagc 25
|||||
Db 173 GTTGCTGGATCTGCTGTTGAAGCG 197

RESULT 11
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS Sequence 6 from Patent EP0798384.
DEFINITION

```

```

LOCUS       AR101809      5872 bp      DNA
DEFINITION  Sequence 1 from patent US 6083712.
ACCESSION  AR101809
VERSION    AR101809.1  GI:12812607
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 5872)
AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE     Biotechnological method of producing biotin
JOURNAL   Patent: US 6083712-A 1 04-JUL-2000;
FEATURES   Location/Qualifiers
            source
              1..5872
              /organism="unknown"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctgattctgtgtttgaagcg 25
|||||
Db 173 GTTGCTGGATCTGCTGTTTGAAGCG 197

RESULT 12
LOCUS       AR101810      5872 bp      DNA
DEFINITION  Sequence 6 from patent US 6083712.
ACCESSION  AR101810
VERSION    AR101810.1  GI:12812608
KEYWORDS   .
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 5872)
AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE     Biotechnological method of producing biotin
JOURNAL   Patent: US 6083712-A 6 04-JUL-2000;
FEATURES   Location/Qualifiers
            source
              1..5872
              /organism="unknown"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctgattctgtgtttgaagcg 25
|||||
Db 173 GTTGCTGGATCTGCTGTTTGAAGCG 197

RESULT 13
LOCUS       AE000180      11022 bp      DNA
DEFINITION  Escherichia coli K12 MG1655 section 70 of 400 of the complete genome.
ACCESSION  AE000180 U00096
VERSION    AE000180.1  GI:1786988
KEYWORDS   .
SOURCE     Escherichia coli K12.
ORGANISM   Escherichia coli K12.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE  1 (bases 1 to 11022)
AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,

```

```

Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
Mau,B. and Shao,Y.
The complete genome sequence of Escherichia coli K-12
Science 277 (5331), 1453-1474 (1997)
97426617
9278503
2 (bases 1 to 11022)
Blattner,F.R.
Direct Submission
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
3 (bases 1 to 11022)
Blattner,F.R.
Direct Submission
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
4 (bases 1 to 11022)
Plunkett,G. III.
Direct Submission
Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using GeneMark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 [e-mail: mark@ember.gatech.edu]. Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli Genetic
Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlyn. A public version of the database is accessible
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
its annotations are periodically updated; this is version M54. No
sequence changes. Annotation updates: updated gene identifications
and products; all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of
labels. This should allow them to be searched for in Entrez as gene
names.

FEATURES
            Location/Qualifiers
            1..11022
            /organism="Escherichia coli K12"
            /strain="K12"
            /sub_strain="MG1655"
            /db_xref="taxon:83333"
            <1..130
            /note="REP (repetitive extragenic palindromic) element;
            contains 4 REP sequences"
            complement(147..1430)
            /gene="ybhC"
            /note="b0772"
            complement(147..1430)
            /gene="ybhC"
            /function="orf; Not classified"
            /note="f427; 98 pct identical to fragment YBHC_ECOLI
            SW:P46130 (300 aa) but contains 127 additional C-terminal
            residues"
            /codon_start=1
            /transl_table=11
            /product="putative pectinesterase"

```

```

/protein_id="AAC73859.1"
/db_xref="GI:1786989"
/translation="MNTFVSRLALALAFGLTILACSTPPDQRPSTQAPGTSRRPI
LSAEQNFDAQHYFSLTGLPAAANSPITLPAQDFVVGPAQGTGVTHTTIOAAVD
AAIILKRNKROYIAVMPEYOGTYVPAAPGGITLYGTGEKPIDVKIGLSDGMSPA
DWRHDVNPGRKMPKPAWYDSCQSRSDIGVLCSAVFSQNNGLQLOLNTIENT
LGDSDVAGNHFAVLRDQGVQINNVLGRONTFFVTNSGVQNRLETNRPRTLV
NSYIEGDVDIVGARGAVFONTEFRVNSSTQGEAYFAPATLSNIYYGLFVNSRFD
AFGQVAGLGRSLDVANTNGQVIRDSAINEGFNTAKPWADAVISNRPFGNTGSVD
DNDEIQRLNDINTNRWENNNRGVSGSKVAEAKK"
/notes="factor Sigma70; predicted +1 start at 806574"
/complement(1507..1535)
/notes="ybhA"
/notes="ybhA"
/notes="ybhA"
/complement(1582..2058)
/notes="ybhA"
/complement(1582..2058)
/notes="ybhA"
/function="orf; Unknown"
/notes="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC73860.1"
/db_xref="GI:1786990"
/translation="MKLISNDLRDQGLKPHRHVFMGVDGDNISPLHAWDDVPAGTK
SFVVCYDPDAPGSGWHVNVNLPADTRVLPQFGSGILVAMPDGLQTRTDFGTG
YDGAAPKGETHYIFTVHALDIERIDVDEGASGMVGNVHFHSLASASITAMFS"
/complement(2108..2124)
/notes="central position to predicted promoter:85.5"
/bound_moiety="Rhas predicted site"
/complement(2117..3406)
/notes="bioA"
/notes="b0774"
/complement(2117..3406)
/notes="bioA"
/EC_number="2.6.1.62"
/function="enzyme; Biosynthesis of cofactors, carriers:
Biotin"
/notes="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
/codon_start=1
/transl_table=11
/product="7,8-diaminopelargonic acid synthetase"
/protein_id="AAC73861.1"
/db_xref="GI:1786991"
/translation="MTDLDLFDQRIWHIHYPTSMTPLPVPVVSAGCELILSDGRR
LVDGMSWAAIHGYNHNPOLNAKSKIDAMSHVMEFGGTHAPALCRLKLVAMPQ
LECVFLADSSVAVEVAMKALQYQAKGEARORLFRNGYHGTDFGAMSVCDPDS
MHSUWGYLPENLFAPOQSDGWEDEMDVGFARLMAAHRHEITAIVIEPIVOCAG
GMRYHPEWLKRIKICDREGILLIADEIATGFGRTGKLFACHEAETAPDILCKGAL
TGTMTLSATLTTRVAVETISNGEAGCFMHPGTFMGNPLACAAANASLAILESGDWOQ
QVADIEVQLREQLAPARDAEMVAVRVLGAIGVVEVTHPVNMAALQKFFVEQGVLRP
FGKLIYLPPIIILPQQLRLTAAVNRVQDETFCCQ"
/complement(2193..2221)
/notes="bioA"
/notes="factor Sigma70; predicted +1 start at 807260"
3411..3450
/notes="central position to bioB promoter: -20"
/bound_moiety="BioB documented site"
3411..3450
/notes="central position to predicted promoter: -20; genetic
evidence for the site"
/bound_moiety="BioB predicted site"
complement(3411..3450)
/notes="central position to predicted promoter:50"
/bound_moiety="BioB predicted site"
complement(3411..3450)
/notes="central position to bioA promoter:50"
/bound_moiety="BioB documented site"
3413..3441
/notes="factor Sigma70; promoter bioB; documented +1 at
808525"
complement(3447..3473)
/notes="factor Sigma70; promoter bioA; documented +1 at

```

```

808515"
3493..4533
/genes="bioB"
/notes="b0775"
3493..4533
/genes="bioB"
/EC_number="2.8.1.-"
/function="enzyme; Biosynthesis of cofactors, carriers:
Biotin"
/notes="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
/codon_start=1
/transl_table=11
/product="biotin synthesis, sulfur insertion?"
/protein_id="AAC73862.1"
/db_xref="GI:1786992"
/translation="MAHRPRWTLTSOVTELFKPLDLDLLEFAOQVHROHDFDPROVOVST
LISIKTGACPECKYCKPOSSRYKTGLEARLMEVQVLESARKAKAAGSTRFCMGAW
KNPHERDMPYLEOMVQGVKAMGLEACMTLGTLSAQOARLANAGLDIYNHNLDTSPF
YGNITTRYIQRDLTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLQLIANLTPPE
VFINMLVKGFTPLADNDVDADFRTIIVARIIMP"SYRLSAGREQMNPQTQAMC
FMAGANSIFYGCKLLTTPNPEEDKDLFRKLGLNPOQTAVLAGDNEQOQRLQALMT
PDTDEYNAAL"
4530..5684
/genes="bioF"
/notes="b0776"
4530..5684
/genes="bioF"
/EC_number="2.3.1.47"
/function="enzyme; Biosynthesis of cofactors, carriers:
Biotin"
/notes="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
/codon_start=1
/transl_table=11
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAC73863.1"
/db_xref="GI:1786993"
/translation="MSMQEKINAAALDARRADALRRRYPVAGQAGRWLVADROYLNF
SSNDYLGISHHPQIIRAOQQAEGFGSGGSHVSVVHQALESEALMGYSRA
LLFISGFANQAVIAAMAKEDRIAADRLEASHLLEASLSQSLRRFAHNDVTHLAR
LLASPCQGVVTEGVFSDGDSAPLAEIOQTQOHNGWLMVDHAGTGVIGQGRG
SCWLQKVRPELLVTFGRGFGVSGAAVLCSSVADYLLQFAHLLIYSTSMPPAQAL
Query Match 100.0%; Score 25; DB 1; Length 11022;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gttgtggtatctgctgtttgaagcg 25
|||||
Db 3549 GTTGCTGGATCTGCTGTTGAAGCG 3573
RESULT 14
AE005258 13501 bp DNA BCT 21-MAR-2001
LOCUS Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
of 155.
DEFINITION
ACCESSION AE005258
VERSION AE005174
KEYWORDS AE005258.1 GI:12513751
ORGANISM Escherichia coli O157:H7 EDL933.
Escherichia coli O157:H7 EDL933
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 13501)
AUTHORS Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,
Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
Posfai, G., Hackett, J., Link, S., Boutin, A., Shao, Y., Miller, L.,
Grotbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamou, K.,
Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
Weich, R.A. and Blattner, F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
JOURNAL Nature 409 (6819), 529-533 (2001)
MEDLINE 21074935

```

11206551  
 2 (bases 1 to 13501)  
 Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,  
 Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,  
 Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,  
 Grobeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K.,  
 Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,  
 Welch, R.A., and Blattner, F.R.  
 Direct Submission  
 Submitted (22-OCT-2000) Laboratory of Genetics, University of  
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  
 FEATURES  
 source  
 1. .13501  
 /organism="Escherichia coli O157:H7 EDL933"  
 /strain="EDL933"  
 /serotype="O157:H7"  
 /db\_xref="taxon:155864"  
 /note="enterohemorrhagic"  
 <1. .7576  
 /note="O-island #36; Region of the EDL933 chromosome not  
 homologous to E. coli K-12 MG1655; Cryptic prophage  
 Cp-933k; includes one copy of the 13 bp direct repeat that  
 flanks the prophage"  
 66. .665  
 /gene="lomk"  
 /note="lomk"  
 /note="z0981"  
 66. .665  
 /gene="lomk"  
 /function="putative membrane; Other or unknown (Phage or  
 Prophage Related)"  
 /note="Residues 1 to 199 of 199 are 72.86 pct identical to  
 residues 1 to 199 of 199 from GenPept 118 ;  
 gi|7532789|gb|AAE63231.1|AF151091\_2 (AF151091) Lom  
 [prophage p-Eiba]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative outer membrane protein of prophage  
 Cp-933k"  
 /protein\_id="AAG55137.1"  
 /db\_xref="GI:12513752"  
 /translation="MRKVCAAILSAALCLAVSGVPWASEHQSTLSAGYLHASTDAPG  
 SDDLNGINVRYEFTDLGLITFSYANAEDOKTHYSIDRWEDYVRNWFVSNVAG  
 PSVRNWFSAVAGVAYSRVTSFGDYEPRVTDNKRKTHDVLGTGDDARYSNTSLAW  
 GAGVQNPTEAVAVAYEXSGSGDWRDTEFIVGVGYKF"  
 719. .2041  
 /gene="z0982"  
 719. .2041  
 /gene="z0982"  
 /function="putative structure; Structural component (Phage  
 or Prophage Related)"  
 /note="Residues 164 to 440 of 440 are 68.79 pct identical  
 to residues 381 to 645 of 645 from GenPept 118 ;  
 gi|45854361|gb|AAD25464.1|AF125520\_59 (AF125520) putative  
 tail fiber protein [Bacteriophage 933W]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative tail component of prophage CP-933k"  
 /protein\_id="AAG55138.1"  
 /db\_xref="GI:12513753"  
 /translation="MGMAVQISGLVKDAGKPIQNCITQLKARNSTXVVNVTASE  
 NPDEAGRYSMDEVGYQSVTLVVEGFFPSHAGTISVYEDSQPGTLNDFLAGMTEDDAR  
 PEALRRFPQVEEARHAEAKKNAGEAETSARNAGISASKAEASANADTSAEDASE  
 SARQAEASAKKSEASEASEAKQAKSESQSATDAELSKKTAESAAGNAARDA  
 TSTKAKRESAESQASBQSRIAEADAVNRITVYVPGPKGPGAGPKGPKDGE  
 RGDTPAGATGERGGDTGPGQPGKDGGERGETGLTGNAGPGPKGDTGAAGPA  
 GPQPGKGTGAAGTGPQPGPKDGTGQIRFLGPMRIETNSYGFPGTGDGALI  
 TGLTFLDPKDATQVQGMFQHLQVRFGDGPWQDVRGLDEVGSDTGTGTE"  
 2004. .2312  
 /gene="z0984"  
 2004. .2312  
 /gene="z0984"  
 /function="orf; Other or unknown (Phage or Prophage  
 Related)"  
 gene  
 CDS  
 /note="Residues 1 to 102 of 102 are 98.03 pct identical to  
 residues 28 to 129 of 129 from GenPept 118 ;  
 gi|4585437|gb|AAD25465.1|AF125520\_60 (AF125520)  
 hypothetical protein [Bacteriophage 933W]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown protein encoded by prophage CP-933k"  
 /protein\_id="AAG55139.1"  
 /db\_xref="GI:12513754"  
 /translation="MKWAVIOAELENDNMILRKLMSLCGCGKHKHDCENGSRSLTAQLR  
 LGPADILESDENGIIPEQDRVITQVQVILDADKKQIQCVVRPLQILRADGTWENIGGMK  
 "  
 2489. .3469  
 /gene="z0985"  
 2489. .3469  
 /gene="z0985"  
 /function="orf; Other or unknown (Phage or Prophage  
 Related)"  
 /note="Residues 13 to 326 of 326 are 55.73 pct identical  
 to residues 18 to 331 of 336 from GenPept 118 ;  
 gi|6960367|gb|AAF33527.1| (AF170176) hypothetical protein  
 predicted by Glimmer [Salmonella typhimurium LT2]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown protein encoded by prophage CP-933k"  
 /protein\_id="AAG55140.1"  
 /db\_xref="GI:12513755"  
 /translation="MLSPIRTTFHNSVINIVSSPQTVSFAGKEYELKVIDEKTPILF  
 QWFENPERYKKDEVPIVNTKQHPVNTAARTIESDMIGIFVDFGFSVNMKTAFS  
 ELSKQDLSLIMYAKRNHAWDFEDLNALLKAGEIFRCTYNTKHGIFSFGCGLYLD  
 NMMLITGLGTIYAPDGLISMHVDRRNSVNTENSALIVNRSNHPALLEGLSPMHKSDV  
 AHPYDGLGKGVKKYFNFTPLNHNHFCDFIEFNHPNIMNTSQTCTSSW"  
 3530. .4522  
 /gene="z0986"  
 3530. .4522  
 /gene="z0986"  
 /function="orf; Other or unknown (Phage or Prophage  
 Related)"  
 /note="No significant matches"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown protein encoded by prophage CP-933k"  
 /protein\_id="AAG55141.1"  
 /db\_xref="GI:12513756"  
 /translation="MKIPSLQSNFNSAPAGYSAPIAPNRAENAYADVLDIGKRPL  
 SAADLSNVYESVIRAVHDSRLIDQSLTQVDMIGNTVLDLRSQIFRDAVSYGHNK  
 VHIGTKRYNEVELNEESSVKIDDITQSLTCELYEDVQCEPIFFICEAGENDNEEPY  
 VSFVAPDFTDSYEMPSWQGLTHEIHHVTGSSDPSGDSNIELGTEILARRAQELG  
 WSPVPDKGYAEPEREAHLRLNLALRQAMRHEBERAFERLGTISDRYBASPDT  
 EYSVSNICGYGFIQQHDFPGLAINDLQDANQILYHGAPYIFTFGDVKHNOQ"  
 5350. .6231  
 /gene="z0989"  
 5350. .6231  
 /gene="z0989"  
 /function="orf; Other or unknown (Phage or Prophage  
 Related)"  
 /note="No significant matches"  
 /codon\_start=1  
 /transl\_table=11  
 /product="unknown protein encoded by prophage CP-933k"  
 /protein\_id="AAG55142.1"  
 /db\_xref="GI:12513757"  
 /translation="MLSPVSNLGGCSNLSLTRLTSPONRVLSVVRDAVHSDNGAQQ  
 KRGIVSRVATDNKFCVTRSHSGCTNLLRLGWPKGEISRKLEVLNAGSPVSAAM  
 EGVIVSNRDLTPVDYAPPELPSVDYNLSVPGNVGKGVNAVVEDADATKVLKM  
 FTISQNEEDTSEVRCFNQIYAGSAGEKIYGNNGDILIGIMDKINGESLILNLSLPAQ  
 AEHAIYDMEDFRLQKGLIFDVTETNVLNDRKNEFNIDISYNVNSDRSRESQIMQ  
 SYHGGKQDLISVWLSK1"  
 6462. .7160  
 /gene="z0990"  
 6462. .7160  
 /gene="z0990"

```

/function="orf; Other or unknown (Phage or Prophage
Related)"
/Note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55143.1"
/db_xref="GI:12513758"
/translation="MRPISLNLVHQSSRSSMSDIESLVKASSVQWIKNNPQLRF
QGTDNHITQQLEALDKIGTETGRVLLNAIESISRLKSETVVIHLSSRLGVMAHRD
IDAENHRGTSGDFHCNNAVEYPCGEGISVVDHFATIVFHELLHVFILNGLRKLKVS
SRAESQKYSPLLEEARTVGLGAFSEEVLSNKKFEEIGMPRTSYDPXDSALIHDNT
VSLGFQQVRLHPLI"
/complement(7668..8144)
/gene="ybha"
/Note="Z0992"
/complement(7668..8144)
/gene="ybha"
/Note="Z0992"
/complement(7668..8144)
/gene="ybha"
/function="orf; Unknown function"
/Note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="MKLISDLRDCDKLPHRVNMGYDGDNISPLHAWDDVPAGPK
SFVYCDPDADPTGSGMHWVNLPAIDRVLPQGFSGVLVAMPDGLQRTDFGKTG
YDGAAPKGETHRYIFTVHALDVERIDVDEGASGMVGFNVHFHSLASASITAMFS"
/complement(8203..9492)
/gene="bioA"
/Note="Z0993"
/complement(8203..9492)
CDS
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gttgtgagatcgtctgtttgaagcg 25
|||||
Db 9635 GTTGCTGGATCTGCTTTGAAGCG 9659

RESULT 15
AP002553 297816 bp DNA BCT 07-MAR-2001
LOCUS
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION AP002553 BA000007
VERSION AP002553.1 GI:13360211
KEYWORDS
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (sites)
AUTHORS Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Hayashi,T., Honda,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C., and Shinagawa,H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780
REFERENCE 2 (sites)
AUTHORS Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655

```

---

```

JOURNAL MEDLINE REFERENCE
3 (sites)
Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Shinagawa,H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
JOURNAL MEDLINE REFERENCE
4 (sites)
Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
JOURNAL MEDLINE REFERENCE
5 (bases 1 to 297816)
Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
Hayashi,T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail:Ken@gen-info.osaka-u.ac.jp,
URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
Fax:81-6-6879-2047)
genome project:
Location/Qualifiers
1..297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
79..1245
/gene="ECs0753"
/gene="ECs0753"
/Note="similar to SUCC_ECOLI gi11786948 percent identity
100 in 388 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="succinyl-CoA synthetase beta subunit"
/protein_id="BAB34176.1"
/db_xref="GI:13360212"
/translation="MNLHEYOAKQLFARYGLPAPVGYACTTPREAEAAASKTGAPWV
VKQVHAGRGKAGGVKVVNSKEDIKRAFNELGKRLVYQTDANGOPVNLIVRAAT
DIARELYLGAVDSSRRVVFMAETEGVEIEKVAETPHLHKVALDPLTGPMPYQG
RELAKFLGLEGLVQQTIFMGLATIFLERDLATIEI NPLVITKQGDICLDGKLA
DGNALFRQDRLREMQSDQDPREAOAGNELNYALDGNICGVNAGLAWGTWDIV
KLHGEPANFLDVGCGATKERVTEAEKIIISDDKVKAVLNIFGGIVRCDLIADGII
AVAEVGVNVPVVRLEGNAELGAKKLADSLNIIAAGLTDAQOVVAVEGK"
1245..2114
/gene="ECs0754"
/gene="ECs0754"
/Note="similar to SUCC_ECOLI gi11786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"

```

```

AA1YKHEVTPVGVYIAGVTAPKGRMGHAGAIAGGKGTADKFAALEAAGVKTVRS
LAD1GEALKTVLK"
2781. 3686
/gene="ECs0755"
2781. 3686
/gene="ECs0755"
/notes="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator) -
Escherichia coli gi1417043|spiP32064|GCV_A_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/translation="MRGKIPKTELLVTFFVVARHESYTRAEELATQSAVFRQVSAL
EEFLRPLFHSKRIPLNDTGYIGYIVKTEKLNKLRDPTINTIMTWOPTVQVYELAVN
PTFSTHMLIPLNHEFTKLHPDIIVNIHSLANNDFLNREYDAVIMRENFCAPWAEVEY
LFEEIILPVCSSGLLMSDOKLSVAELLTEPLLHQSTRTITGWEWFALSGVSSPLVN
NGPFDLLSMLIAAVRSNLGVALLPFAIQHDLSDGMWPCDVPRTGNRFTMTWQE
EKSDSPHLQFREWLLAKSVVQEM"
complement(3720. 4322)
/gene="ECs0756"
complement(3720. 4322)
/gene="ECs0756"
/notes="probable cob(I)alamin adenosyltransferase, similar
to cob(I)alamin adenosyltransferases (corrinoid
adenosyltransferases) e.g. [Escherichia coli]
gi115148|spiP13040|BTUR_ECOLI percent identity 67 in 200
aa"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="putative cob(I)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARISTRHVRQOQKLEQVDTRVAAAEKKGILIVFTGNGKG
KSTAAFTGTVRAGCHTKGVQAYIRKQDNGEYNLQPLGVFEHMGTFGWETQNR
KADIDAAKEVWSKSKMLADKDYLDVLDLTYMLAYHLDTEVIASLQNRPAQOSV
IVTGRGCHSOLIKMADTVSEIRPVKHAFDNGIQAPGIDW"
complement(4332. 5984)
/gene="ECs0757"
complement(4332. 5984)
/gene="ECs0757"
/notes="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarase) - Escherichia coli
gi1120598|spiP00923|FUMA_ECOLI percent identity 68 in 545
aa"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"
/translation="MSKPFVQEPFLQNKDCTEYTLISQHIITVTELDGEVVIKIAPE
ALTLLARQAYEASFURSARHLOQVASILNDQOASNDKTVALLQLRNAEVSAGKVLV
NCDDTGTATIVASGQOIWTGGNDAAESKGIYSTFOENNRFSQNPALDMDYEVNTQ
TNLPAQDIDSAVAGDEYHFLCVNKGGSANKAALYQETKSLIQPERKTLFLIEKMSL
GTACPPYHIAFVVGGLSADQTLKIAKLASTKYVDNLPTSGNEQGAFRDIELEKVL
EASQFCGIAQGGKGYFAHDIRVIRLPRHGGSCPIMALSCSADRNKAKINKHGTLW
EKLEHNPQGVIPASLREENAHQVLDLNPRLPDMODLARLPVGRVSLSGPIVVAR
DIAHAKIKARLDGSEMPKYLKHIIYIYAGPAKTPENMACGSLGPTTGGRMGDIITF
GAAGSLVLSKGRNSQOQVTDACHKGGFNLSIGGAALLAOEYVKSRLCRLLEYPELG
MEAVMMVEENLPAFLVDKGNFFSQFQQHRCASCAPAG"
complement(6092. 7372)
/gene="ECs0758"
complement(6092. 7372)
/gene="ECs0758"
/notes="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate

```

```

symport proteins) e.g. [Bacillus stearothermophilus]
gi121467|sp|P24943|GLTT_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transpor"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKKISLTMTMILLALVGLMIGVLLNNTASPETAKLYAEISIFT
TIFLRLMIAPLVVSTLVVGIAGKMGDAKALGRIFSKTLEFICASLLSIALGLITY
NFFMPGTGINFVAHGAFTGVAAEPFTLKVFISHAFTSIVDAMAHNEIIQIVVFSI
FIGLSLTAIGFSGSAIVHALDSLAHMLKLTGYVMLFAPLTFVFAISALIAERGLAVM
VSAGIFMGFEYFTMLLWVLLIGLAIIVVGPICIRRLRALSEPALLAFTSSSEAAFP
GTLEKLEGFVSPKIASFVLPIGYSFNLVGSMAYCSFATVIAQACNHLHSIGQITM
LLILMLTSGKNAGVPRAVMVIAATLNQFNPEAGLLMLMGVDFPLDMGRSATNVMNS
AMGAMVSRWEGEHGEGECRGKALKPNESNALP"
complement(7533. 7850)
/gene="ECs0759"
complement(7533. 7850)
/gene="ECs0759"
/notes="unknown"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"
Query Match 100.08; Score 25; DB 2; Length 297816;
Best Local Similarity 100.08; Pred. No. 0.42; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0;
QY 1 gttgtggtatctgtgtttgaagcg 25
|||||
Db 92367 GTTGCTGGATCTGCTGTTGAAGCG 92391
|||||
RESULT 16
E00893 1121 bp DNA PAT 29-SEP-1997
LOCUS Genomic DNA encoding biotin Synthetase.
DEFINITION E00893
ACCESSION E00893
VERSION E00893.1 GI:2169154
KEYWORDS JP 1986149091-A/1.
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1121)
AUTHORS Hirono, Y., Kojima, T. and Kimura, H.
TITLE DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
JOURNAL PRODUCTION OF BIOTIN
COMMENT Patent: JP 1986149091-A 1 07-JUL-1986;
NIPPON SODA CO LTD
OS Escherichia coli
PN JP 1986149091-A/1
PD 07-JUL-1986
PF 24-DEC-1984 JP 1984272605
PI HIRONO YOSHIIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
C12N15/00, C12N1/20, C12P13/18, (C12N1/20, C12P13/18), (C12P13/18, PC
C12R1:19);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: strain=Escherichia coli Nsl01;
CC Feature is identified by experimental;
FH key Location/Qualifiers
FT CDS 42..1079
FT /product="biotin synthetase".
FT Location/Qualifiers
1..1121
FEATURES
source

```



```

/organism="Escherichia coli"
/db_xref="taxon:562"
289 a 296 c 325 g 211 t
ORIGIN

Query Match 96.0%; Score 24; DB 10; Length 1121;
Best Local Similarity 100.0%; Pred. No. 0.82;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgctgagatcgtctgtttgaagc 24
|||||
Db 98 GTTGCTGGATCTGCTTTTGAAGC 121

RESULT 17
AE004192 12891 bp DNA BCT 31-JUL-2000
LOCUS
DEFINITION
Vibrio cholerae chromosome I, section 100 of 251 of the complete
chromosome.
ACCESSION AE004192 AE003852
VERSION AE004192.1 GI:9655581
KEYWORDS
SOURCE
Vibrio cholerae.
Vibrio cholerae
Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
1 (bases 1 to 12891)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
Sellers, P., McDonald, L., Utterback, T., Fleishmann, R.D.,
Nierman, W.C. and White, O.
DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae
Nature 406 (6795), 477-483 (2000)
2 (bases 1 to 12891)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
Sellers, P., McDonald, L., Utterback, T., Fleishmann, R.D.,
Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,
Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
Direct Submission
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers
1..12891
/organism="Vibrio cholerae"
/strain="N16961"
/serotype="O1"
/db_xref="taxon:666"
/chromosome="I"
/feature="biotype: El Tor"
complement(84..1370)
/gene="VC1111"
complement(84..1370)
/gene="VC1111"
/feature="similar to GB:J04423 SP:P12995 PID:457106
PID:490217 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="adenosylmethionine-8-amino-7-oxononanoate
aminotransferase"
/protein_id="AAF94270.1"
/db_xref="GI:9655582"
/translation="MMDLAFDRCHTHWPTYSLTPLTCTPVSRADGVMTLEDGRSL
VDGSSWAAIHGHPKLNQAAHQIQEOMSHIMFGLTHQPAIELCQKLLKAPNNL
QQVFLADSGSVAEVLKMAQYHKSQPRKFLTLRHGYHGDTEAAMSVDPDNSM
HSLYKGLPEHIFANSPEGGFDAWDERDIADFRHKLTEHHHQAIAAVILEPIVOGAGG
MRLYHPEFLROYRALCDERGVLLILDEIATGRTGCKLFPACRHAGIOPDILCVGKALT
GGYMTLSALTTQOVADTVCAGAGCFMHGPTFMGNPLACAVASLSLIEQDQWQQ
VANIEAFFAPQLKPLSDPRVKOTRWLGAVGVVETHPLVMNEVIALFVEHGVWIRPE
GRLIYLMPPYISQPEHLQKLVDAAMAALNQPDGFKQ"
1497..2549
/gene="VC1112"
1497..2549
/gene="VC1112"
/feature="similar to GB:J04423 SP:P12996 PID:145425
PID:490219 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="biotin synthase"
/protein_id="AAF94271.1"
/db_xref="GI:9655583"
/translation="MEVRHNTVAEYKALLDKPFMDLFEAQVVRHRRHHPNHVQVST
LISIKTGACPECKYCPQSAHYRTDVKRLMEVERVLDAQAQKNSGSTRFCMGAAW
KNPKRDMPLKEMIRGVKDMGLTGMTLTPOAQOLAQGLDYNNHLDTSPEF
YGNILTRYQDLPLSHVRDAGMKICSGGLIGGESTNDRAGLLVELANLTPHES
VPINMLYKVGPIPEQVDVDFVRLIAVARIMPKSAVRLSAGRKMEQMALC
FMAGANSIFYGCKLLTPPAEDSDMLLPKLGINNEQVAQRPDEITENELLDVRVER
VAARPTASDLFYDAAL"
2533..3687
/gene="VC1113"
2533..3687
/gene="VC1113"
/feature="similar to GB:J04423 SP:P12998 PID:145426
PID:490225 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAF94272.1"
/db_xref="GI:9655584"
/translation="MTLRFKSRTEQALNREHSRLTRRLQLLDRSAQGLIQODRTYL
NFSSNDYLGANDAEVLQAWOTCLARFGAGSGSPMTGFSAAHAELHELCHMLGYE
RALLVSSGFSANQALLFTLLEKEDLLQKLNHASIMEAGMLSPATMRFRKNDTEHL
RQLLHQSNLSVTEGVFSDGDCAPLAQIATLVKORAWLMDVDDAHGVLGEGAGS
CQAAGTHPELVVTFKGAGLAAVLCDAHVGDYLTQFAHHVYSTAMPQAQHALT
HAARMIQSQSWRREQLAELLACFDEQCRNIPGVATQTIKPLIGSSDSALSASHEL
KQGIWVSARIPPTVPVGSARLITITTAHTQAQIRLSEQLKHVMEKDT"
3549..4487
/gene="VC1114"
3549..4487
/gene="VC1114"
/feature="similar to GB:J04423 SP:P12999 PID:145427
PID:490221 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="biotin synthesis protein Bioc"
/protein_id="AAF94273.1"
/db_xref="GI:9655585"
/translation="MCECHSTTHRACQCAFTYHDNGGAYPSADTPLIGAVKTCDEG
HMSMTATELCVELKDKSAIAQAGKAAHYDQHAARQDVGRLILQKRPSCGLGLRVL
DLGCGTGYFSALLRERGAVVCAVADISHAMEQAKORCDEGMSYQLADAPFPASAC
FDMVFSSALQWCEDELSPLSEIRRVLPKHGAFSLTLDGSLFLEQAQWRSVDHHRH
INQFISINQVIALAQAGCSQHLLAAITVWYEAFFMLMRDLKLGIGANHVSGRSTGL
ISRRTLAKVELAYQSFRRNQGLVPATYQVCLGVTHR"
4463..5179
/gene="VC1115"
4463..5179
/gene="VC1115"
/feature="similar to GB:S68059 SP:P13000 PID:490223
PID:145428 PID:460901; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase"
/protein_id="AAF94274.1"
/db_xref="GI:9655586"
/translation="MFRGYSRMRNAIFTAGTDDTGVKTVASKAIIQALATHNIATIGY

```



DTB synthetase; KAPA synthetase; biotin operon; biotin synthetase; dethiobiotin synthetase.  
 Serratia marcescens (strain: Sr41) DNA.  
 Serratia marcescens  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Serratia.  
 1 (bases 1 to 7215)  
 Sakurai, N.  
 Direct Submission  
 Submitted (25-AUG-1993) to the DDBJ/EMBL/GenBank databases. Naoki Sakurai, Tanabe Seiyaku Co., Ltd., Res Lab of Applied Biochemistry; 2-50, Kawagishi-2-chome, Toda-shi, Saitama 335, Japan  
 (E-mail: GFC01101@niftyserve.or.jp, nsakurai@ddbj.nig.ac.jp, Tel: 048-433-2545, Fax: 048-433-2540)  
 2 (bases 1 to 7215)  
 Sakurai, N., Inai, Y., Akatsuka, H., Kawai, E., Komatsubara, S. and Tosa, T.  
 Complete nucleotide sequence of biotin operon of Serratia marcescens  
 Unpublished (1993)  
 Submitted (25-Aug-1993) to DDBJ by:  
 Naoki Sakurai  
 Res. Lab. of Applied Biochemistry  
 Tanabe Seiyaku Co., Ltd.  
 2-50 Kawagishi-2-chome  
 Toda, Saitama 335  
 Japan  
 Phone: 048-433-2545  
 Email: nsakurai@ddbj.nig.ac.jp  
 Fax: 048-433-2540.  
 Location/Qualifiers  
 1..7215  
 /organism="Serratia marcescens"  
 /strain="Sr41"  
 /db\_xref="taxon:615"  
 complement(1353..2630)  
 /gene="bioA"  
 complement(1353..2630)  
 /gene="bioA"  
 /EC\_number="2.6.1.62"  
 /codon\_start=1  
 /transl\_table=11  
 /product="7, 8-diamino-pelargonic acid aminotransferase (DAPA aminotransferase)"  
 /protein\_id="BAA04284.1"  
 /db\_xref="GI:402531"  
 /translation="MSVTASDLAFQDHIIHWPYTSMSRPLPCYPPIESAGSVELQLADG RSLVDMSSWAAIHGYNHPRNLQAQOOLKMSHMFGGITHPAALSLCRLRVEMTP EALQCVFLADSGVAVEVSLKMALOYWOARGERRORILTLRHGYHGDTFGAMSVCDPD NSMHSLYOCYLAPHLPATAPOCDFEWEEDLAPPALEOHAGEVAALVLEPVVQG AGCMRIYHTYLRVRDVYVAIKLLLLADEIATGFGRTGKLFACHEAQVVPDILCLG KALTGYMTLSALTTRHVAETISNGAAGCFMHGPTFMGNPLACAVADASLALLAENR WQOVSAIETQLKRELLPLGRWRPHVRVLGAIGVEMPEPVDVAELQRGFVERGVW IRPFGKLIIMPYPYITEAEQLSRLTAAVADAAR"  
 2716..3756  
 /gene="bioB"  
 2716..3756  
 /gene="bioB"  
 /codon\_start=1  
 /transl\_table=11  
 /product="biotin synthetase"  
 /protein\_id="BAA04285.1"  
 /db\_xref="GI:402532"  
 /translation="MMADRIHWTVQQAQALFDKPLLELLFEAQTVHRQHFDPRQVQVS TLLSIKTGACPECKYQPSQRYKGTGLSESLRMQVEVLESARKANGSTRFCMGAA WKNPHRMDPYQVQVQVGMKMETCMTLTGLDGTQAEALAEAGLDYNNHNLDSPE FYGSIITTSYQERLDTLKDVGAKIVCSGIGVLTGRTVDRAGLLVOLANLPKPE SVPINMLVKVGTPLADNDVDPDFPIRTIAVARIMPPSSVRLSAGREOMNEQTQAM CFNAGANSIFYGCKLLTTPNPEDKDLQLFRKLGLNPQQTATEHGDNQOQVVLAKQLL NADTAEFYNAAP"  
 3756..4904  
 /gene="bioF"  
 3756..4904  
 /gene="bioF"

/gene="bioF"  
 /codon\_start=1  
 /transl\_table=11  
 /product="7-keto-8-aminopelargonic acid synthetase (KAPA synthetase)"  
 /protein\_id="BAA04286.1"  
 /db\_xref="GI:402533"  
 /translation="MSWQRIEQALAEERRLNAAYRRRTOTTEGGNGRQIRLGRDLNLF SGNDYLGISODARVIAAWOAGARYGVNGSGGHVTGFSAAHOALFEOLAAMGLCPRAL LFTSYAANQAVLAALMOKGBRILADRLSHASLLEACAGAVARRAPVPAQSTAGLARIC LAKPCDQORLAVTEGLFMDGAPLAELHRLTRAAGAWLMVDDHAGVYGGQGRGS CWQOQREGDLVATFGKAFGVGAALDEATAEYLLOFARHLIYSDHAGVYGGQGRGS AALARIREGDLRARLQDNIRFRQGAAPLALTLDSDTAIQPLLLVDGNORALDLATR LREGGLWSAIRPPTVPPGGARLRLLLTAAHQSDIDRLLEVLNDVQ"  
 4888..5655  
 /gene="bioC"  
 4888..5655  
 /gene="bioC"  
 /function="synthesis of pimeloyl-CoA"  
 /codon\_start=1  
 /transl\_table=11  
 /product="the product of bioC"  
 /protein\_id="BAA04287.1"  
 /db\_xref="GI:402534"  
 /translation="MTSANDTVNKOAVASAFSRAAGSYDAAALQORDVGERLLMGSS HPGEQLLDAGCGTGYFSRMEERCKRVTALDLPAGMLDVARQROAAHHYLLGDIEQVP LPDAAMDICFSSLYVQWCSLPAALAEIYRVTRPGGVILFSTLAAGSLQELGDWAOQV DGERHNAFLPLTOITACAAAYRHVELTFLNYPDVTMLMRSLAGIGATHLHQGRE GGLMSRGLAALAAQAYPCRGQFPLSYHLAYGVYRE"  
 5648..6331  
 /gene="bioD"  
 5648..6331  
 /gene="bioD"  
 /function="conversion from 7, 8-diaminopelargonic acid into dethiobiotin"  
 /codon\_start=1  
 /transl\_table=11  
 /product="dethiobiotin synthetase (DTB synthetase)"  
 /protein\_id="BAA04288.1"  
 /db\_xref="GI:912453"  
 /translation="MSKRWFVTGDTVEGKSVASSALLQANRAGYPSAGYKPVASGS EMTAEGLRNGDALALQANSVALDYDEVNPYVFAEPTSPHIVSADEGRPIDAARLSDG LRLRQERADWLVVEGAGGWFTPLSAEYTFADWVROEQLPVILVVGKLGICTINHAVLTA QAVQAGLTLAGWTANDVGGAPGRRHQELATLRMLPRRCWAKSRTCRRPERAPLQGYLDISLQAQ"  
 BASE COUNT 1347 a 2225 c 2367 g 1276 t  
 ORIGIN  
 Query Match 80.8%; Score 20.2; DB 3; Length 7215;  
 Best Local Similarity 88.0%; Pred. No. 63;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Oy 1 gttgctggatctgctgtttgaagcg 25  
 1 ||||| ||||| ||||| |||||  
 Db 2775 GCTGCTGGAACGCTGCTTCGAGCG 2799  
 RESULT 20  
 AF248314  
 LOCUS  
 DEFINITION  
 AF248314 8227 bp DNA BCT 24-JAN-2001  
 Uncultured bacterium pCosAs1 urocanase-like protein (hutU) gene, partial cds; histidine ammonia-lyase-like protein (hutH), DAPA aminotransferase BioA (bioA), biotin synthase BioB (bioB), 7-KAPA synthetase (bioF), biotin biosynthesis Bioc-like protein (bioc), and dethiobiotin synthase Biob (biob) genes, complete cds; ABC transporter-like protein (elsa) gene, partial cds; and unknown gene.  
 ACCESSION AF248314  
 VERSION AF248314.1 GI:12407610  
 KEYWORDS  
 SOURCE uncultured bacterium pCosAs1.  
 ORGANISM uncultured bacterium pCosAs1

```

Bacteria; environmental samples.
1 (bases 1 to 8227)
Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
Direct Cloning from Enrichment Cultures, a Reliable Strategy for
Isolation of Complete Operons and Genes from Microbial Consortia
Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
11133432
2 (bases 1 to 8227)
Entcheva,P., Liebl,W. and Streit,W.R.
Direct Submission
Submitted (21-MAR-2000) Mikrobiologie und Genetik, Universitaet
Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
Location/Qualifiers
1. .8227
/organism="uncultured bacterium pCosAS1"
/db_xref="taxon:143793"
/clone="pCosAS1"
/note="unknown organism, cosmid clone derived from
environmental consortium"
<1. .502
/gene="hutU"
/gene="hutU"
/note="putative urocanate hydratase"
/codon_start=2
/transl_table=11
/product="urocanase-like protein"
/protein_id="AAG53593.1"
/db_xref="GI:12407618"
/transl_table="PRRISFGLPARICWVLEWROKLGAFNEWVRGEVSAPIVI
GRDHLDSGSVASPNRETEAMQDGSVDVSDPPLLNLNTASGATWVSLHGGVGMGF
SOHAGMVICDGTDAARTARVLHNDPATGVMRHADAGYDIAIDCAKEQGLNLPMA
ATQGOH"
502. .2031
/gene="hutH"
502. .2031
/gene="hutH"
/note="putative"
/codon_start=1
/transl_table=11
/product="histidine ammonia-lyase-like protein"
/protein_id="AAG53586.1"
/db_xref="GI:12407611"
/transl_table="MNAALTTPGTLTIAOLRVNQOPLTLDESHEALNDSVACVE
AIVAGRTAGINTGCGLLAQTRIAHDLENLORSIVLSHAAGVGPDDDDIVRLMMV
LKINSLAGSGIKLSVQIALIYNAGVSYDPAKSGVSGASGDPLAPLAHMSUTLLGE
GKARYGEMLPALQKGLAPVTLAAKEGLALLNGTQSTAFALRGLFEADDLFAS
AVVCGALTTEAVILGSRFPDARIHEVRGQIDAAALFRHVLTDTSATAESHHNCDK
VQDPYSLRCPQVMGACLTQMRQVAEVLIVLESNVSDNPLVFAAENEMVFRGNFAEP
VAMAADNLALIAEIGALSERRIALMMDKHMSOLPPLVRENGGVNSGEMIAOVTAAL
ASENKGILCHPHTTSVDKPCPSANQEDHVSMAPAAGRLWEMAGNTRGVIAEWLAAC
QCADRLGLTSSPLLEQARSGCEQVAHYDDDRFFAPDIEAALISLKNKGLSVGLLPAF
L"
complement(2130. .2606)
/note="ORF1; similar to hypothetical 17.1 kDa protein in
modc-bioA intergenic region"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG53587.1"
/db_xref="GI:12407612"
/transl_table="MKLISQDLRDGKPLRHFVNGMGVGDNISPLAWDEVPSGTK
SFVVCYDPPDPTDSGWWHIVANIPADTRVLPGSGSSIALVALPEGAVETRTDFCKAG
YGGAAPKGTHTYITFVHALNVERIEVDEASGAMGVNHFHTLGSATIMAYS"
complement(2651. .3940)
/gene="bioA"
complement(2651. .3940)
/gene="bioA"
/note="7,8-diamino-pelargonic acid aminotransferase"
/codon_start=1
/transl_table=11
/product="DAPA aminotransferase BioA"
/protein_id="AAG53588.1"

```

---

```

/db_xref="GI:12407613"
/transl_table="MTDDLAFDRHOIWHPYTSMTOPLPVPVYVVEAHGCELTASGEQ
LYDGMSSWAAIHGYNHPLNRAAKAQIDRSMHVMFGGITTHQPADYDLRRLVMTPEI
LECVFPAQSGSVAEVAMKMAQYWHAKGESQRFITFRNGTHGDTFGAMSVCDPNS
MISLMKGYLPENLFAPQSRFDGAWDEMVMGLAPLMAAHRHETAAVILFIVQAG
GMRYHPWLKRIIRMCDEGILLIADETATGTRGKLFACFHAGITPDLICLKGAL
TGCTMTSATLTROVAETISNGEACGFMHGPTFMGNPLACAVESLALLSSEGMQD
OVAAESOLOBELAVARSEFVADVRLVGAIGVETTHPVNMAALQRFVARGVWRP
FGKLIYLMPPYVIRPDQLSKLTRAVVDVQAPAHFTR"
4029. .5069
/gene="bioB"
4029. .5069
/gene="bioB"
/codon_start=1
/transl_table=11
/product="biotin synthase BioB"
/protein_id="AAG53589.1"
/db_xref="GI:12407614"
/transl_table="MAHQARWTLTQVTFALFEKPLELLEFQAQIHRQHFDPQIQVST
LLSIKTGACPEDCKYCPQSARYKTGLESERLEMEVEQVLESARQAKNAGSTRFCMGAAM
KNPHERDMFYLEOMVQGVKAMGLQACMTLGTLDHQAORLASAGLDYNNHNLDTPEF
YGNITTRTYOERLDTLDKVRDAGIKVCSGGIVGLCETVTRAGLLLOLANLPTPEF
VPINMLVKVGTPLADNDVAFDXIRTXSVAIXVMXPTXLVRLSTGROMNEOTQAMC
FMAGANSIFYGCKLLTTPNPEDEKDVQLKGLGNPQQTAVLTGDNQEQHLEQQLFN
AUTDQFYNAATV"
5066. .6223
/gene="bioF"
5066. .6223
/gene="bioF"
/note="7-keto-8-amino-pelargonic acid synthetase; similar
to BioF"
/codon_start=1
/transl_table=11
/product="7-KAPA synthetase"
/protein_id="AAG53590.1"
/db_xref="GI:12407615"
/transl_table="MTWQORIDALDARRAADALVRTPVENGAGRFLIRQRRCNF
SCNDYLGISQHPPEIRAWQOQAGYGVGGSGHVSQYTHAHQALEELHAGWLGYPRA
LLFISGFANQALIAALTDKEDRIVADRLSHSLLEAAASHPAQRLRFAHNDVAQLDS
LLDKFCHQQLVVTGIFSMQDSAPLVGIIHAAQROGQSWLLVDDAHGIVTGPGRG
SAAQGVKPELVVTFGKGVGSAALVCSDAVDYLLQAFARHLITSTMPAQAVAL
LAAFNVIRSDENARRORLAALIDOFRAGVRDLPVQITASOSAIQPLIVGNARALHL
AOLRERGIWTAIRPPTVPAGTARLRLTLTAHQPADITLVEALHVAGE"
6207. .6962
/gene="bioC"
6207. .6962
/gene="bioC"
/codon_start=1
/transl_table=11
/product="biotin biosynthesis BioC-like protein"
/protein_id="AAG53591.1"
/db_xref="GI:12407616"
/transl_table="MSQVKNQAIATAAAGFRAQASYSRRDALQORSADGLRALLAENTFA
SYLDAGCGPGNSLAWROTGSVVTAMDLCAPMLEARRNRQAADSYLQADIEAIPLADA
QFSLAWSHLAVQXCASLPRAELAEWRVTQLGKVAFTLLNGSLPELNAQAVDPTOP
HANRTLSCQVNDALSGHQYBCAVQTVTLTFDDLSAMHSLKLGIGATHLHAGRAAKPL
TRGQLQQAQAMPQOAGKFPKLSYQLFHGIIERD"
6940. .7680
/gene="bioD"
6940. .7680
/gene="bioD"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthase BioD"
/protein_id="AAG53592.1"
/db_xref="GI:12407617"
/transl_table="MELSNVTFRVFTGDTFVGKTVASTALLQAAALGRHRTAGYKP
VAGSSEMTAQGLRNSDALALORNSLSALSYPAVNPYTFAPPTSPHYSADQRIEFA
TILSAGLRALTEQADWLVLEGGAGMFTPLSQOQTEADMMWMAERLPVILVGVKGLGINH
AMUQAQTRAGLHFAGVWANGVVPVKKHAEYMATLRVLSAPILGELPWLGGQADT
SAVGQYLDLTALTPTSPSSGQSGAGHPAR"
complement(7619. .8227)
/gene="elsA"

```

## CDS

```

complement(7619..>8227)
/gene="elsa"
/note="putative urea/short-chain amide ABC transporter"
/codon_start=1
/transl_table=11
/product="ABC transporter-like protein"
/protein_id="AAG53594.1"
/db_xref="GI:12407619"
/translation="DCXLRXGXNXLVSCVAGLLPIASGTVMVHEAPGAVNAQGTETI
XVPPDRRIFSLVFNLIHQAKIAGSLSRVFFPELPYLRQRLASGLND
QOYLALANLVTPRLILDEPSGTGLFTVKLAQLLFRNFRGLSLVLLAEQHL
LIRVADRFCLLYGRNVAGRVSELDPLIAHMGPTASQA"
BASE COUNT      1512 a 2642 c 2605 g 1453 t 15 others
ORIGIN

```

Query Match 79.2%; Score 19.8; DB 2; Length 8227;

Best Local Similarity 91.3%; Pred. No. 98;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 tgcgtgatcgtcgtttgaagcg 25

||||| || ||||| ||||| |||||

Db 4087 TGCCTGAAGTCTGTTGAAGCG 4109

## RESULT 21

AC084015/c

LOCUS

DEFINITION Homo sapiens chromosome 10 clone RP11-995B13, WORKING DRAFT

SEQUENCE, 23 unordered pieces.

AC084015

AC084015.9 GI:12039118

HTG: HTGS\_PHASE1; HTGS\_DRAFT.

human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiina; Homo.

1 (bases 1 to 195738)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbara,J.,

Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,

Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn,A.L., Ding,Y., Dinh,H., Douthwaite,K.J., Draper,H.,

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,

Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,

Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,

Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,

Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,

Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,

Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,

Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,

Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,

Louisegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,

Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,

Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,

Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Monabbat,K.,

Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,

Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,

Ogih,M., Okwunou,G., Oragunye,N., Owiedo,R., Pace,A., Payton,B.,

Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,

Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,

Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,

Sisson,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,H.,

Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,

Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,

Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,

Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
Watlington, S., Williams, C., Williamson, A., Wlezyk, R., Wooden, S.,  
Worley, K., Wu, C., Wu, F., Wu, J., Zhou, J., Zorrilla, S., Nelson, D.  
and Gibbs, R.

## TITLE

JOURNAL

REFERENCE

2 (bases 1 to 185738)

Worley, K.C.

Direct Submission

Submitted (09-OCT-2000)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Bay Plaza, Houston, TX 77030, USA

On Jan 5, 2001 this sequence version replaced gi:12007667.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: HBOB

Center clone name: RP11-995B13

----- Summary Statistics

Sequencing vector: M13; L08821

Chemistry: Dye-primer Body: 46% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 166584 bases at least Q40

Consensus quality: 174801 bases at least Q30

Consensus quality: 179020 bases at least Q20

Estimated insert size: 179024; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 3.6x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

consists of 23 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

\* 1 44236: contig of 44236 bp in length

\* 44237 44336: gap of unknown length

\* 44337 63479: contig of 19143 bp in length

\* 63480 63579: gap of unknown length

\* 63580 81827: contig of 18248 bp in length

\* 81828 81927: gap of unknown length

\* 81928 93064: contig of 11137 bp in length

\* 93065 93184: gap of unknown length

\* 93185 105462: contig of 12298 bp in length

\* 105463 105562: gap of unknown length

\* 105563 115530: contig of 9968 bp in length

\* 115531 115630: gap of unknown length

\* 115631 124121: contig of 8491 bp in length

\* 124122 124221: gap of unknown length

\* 124222 133919: contig of 9698 bp in length

\* 133920 134019: gap of unknown length

\* 134020 141345: contig of 7326 bp in length

\* 141346 141445: gap of unknown length

\* 141446 147781: contig of 6336 bp in length

\* 147782 147881: gap of unknown length

\* 147882 153389: contig of 5508 bp in length

\* 153390 153489: gap of unknown length

\* 153490 158140: contig of 4651 bp in length

\* 158141 158240: gap of unknown length

\* 158241 162915: contig of 4674 bp in length

\* 162915 163015: gap of unknown length

\* 163015 166903: contig of 3889 bp in length

\* 166903 167004: gap of unknown length

\* 167004 170365: contig of 3361 bp in length

\* 170365 170464: gap of unknown length

\* 170465 173923: contig of 3459 bp in length  
 \* 173924 174023: gap of unknown length  
 \* 174024 175899: contig of 1876 bp in length  
 \* 175900 175999: gap of unknown length  
 \* 176000 178095: contig of 2096 bp in length  
 \* 178096 178195: gap of unknown length  
 \* 178196 180140: contig of 1945 bp in length  
 \* 180141 180240: gap of unknown length  
 \* 180241 181537: contig of 1297 bp in length  
 \* 181538 181638: gap of unknown length  
 \* 181639 182818: contig of 1181 bp in length  
 \* 182819 182918: gap of unknown length  
 \* 182919 184345: contig of 1427 bp in length  
 \* 184346 184445: gap of unknown length  
 \* 184446 185738: contig of 1293 bp in length.

## FEATURES

source

1..185738  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-995B13"

BASE COUNT 50406 a 41431 c 42341 g 49321 t 2239 others  
 ORIGIN

Query Match 79.28; Score 19.8; DB 76; Length 185738;  
 Best Local Similarity 91.3%; Pred. No. 1.3e+02;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gttgctggtatctgtctgttgaag 23

Db 27828 GTTGTGGAAGTCTGTGGAG 27806

## RESULT 22

AC022397 LOCUS Homo sapiens chromosome 10 clone RP11-428P16, WORKING DRAFT  
 DEFINITION SEQUENCE, 22 unordered pieces.  
 ACCESSION AC022397.4 GI:10881060  
 VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS human.  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 218375)  
 Smith, D.R.  
 Genome Therapeutics Corporation Sequencing Center: Human Genome  
 Sequence Data  
 Unpublished  
 2 (bases 1 to 218375)  
 Smith, D.R.  
 Direct Submission  
 Submitted (03-FEB-2000) Genome Therapeutics Corporation, 100 Beaver  
 Street, Waltham, MA 02453, USA  
 On Oct 19, 2000 this sequence version replaced gi:7549610.  
 ----- Genome Center  
 Center: Genome Therapeutics Corporation  
 Center code: GTC  
 Web site: http://www.genomecorp.com/  
 Contact: gtc-seqcenter@genomecorp.com  
 ----- Project Information  
 Center project name: hg145  
 ----- Summary Statistics  
 Sequencing vector: N/A  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 990315  
 Consensus quality: 197954 bases at least Q40  
 Consensus quality: 206805 bases at least Q30  
 Consensus quality: 208715 bases at least Q20  
 Insert size: 216324; sum-of-contigs  
 Quality coverage: 6.8x in Q20 bases; sum-of-contigs

## FEATURES

source

1..218375  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-428P16"  
 /clone\_lib="RFCI-11"  
 1..1104  
 /note="assembly\_name:Contig19"  
 1205..2615  
 /note="assembly\_name:Contig26"  
 2716..3788  
 /note="assembly\_name:Contig27"  
 3889..5010  
 /note="assembly\_name:Contig28"  
 5111..6325  
 /note="assembly\_name:Contig29"  
 6426..7600  
 /note="assembly\_name:Contig33"  
 7701..8869  
 /note="assembly\_name:Contig39"

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

```

misc_feature      8970..10230
/note="assembly_name:Contig41"
10331..11532
/note="assembly_name:Contig56"
11633..13395
/note="assembly_name:Contig64"
13496..14602
/note="assembly_name:Contig66"
14703..15918
/note="assembly_name:Contig68"
16019..17852
/note="assembly_name:Contig70"
17953..19622
/note="assembly_name:Contig71"
19723..30547
/note="assembly_name:Contig73"
30648..41456
/note="assembly_name:Contig74
clone_end:SP6"
41557..60958
/note="assembly_name:Contig75"
61059..81271
/note="assembly_name:Contig76"
81372..105523
/note="assembly_name:Contig77"
105624..132993
/note="assembly_name:Contig78"
133094..175797
/note="assembly_name:Contig79"
175898..218375
/note="assembly_name:Contig80"
BASE COUNT 60960 a 47360 c 48286 g 59658 t 2111 others
ORIGIN

```

```

Query Match      79.2%; Score 19.8; DB 67; Length 218375;
Best Local Similarity 91.3%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gttgctggatctgctgttgaag 23
||||| ||||| ||||| |||||
Db 129011 GTTGTGGAAGTCTGTTGGAAG 129033

RESULT 23
AC063928
LOCUS
DEFINITION
Homo sapiens chromosome 3 clone RP11-21B16, *** SEQUENCING IN
PROGRESS ***, 75 unordered pieces.
AC063928
VERSION
AC063928.16 GI:10180056
KEYWORDS
HTG; HTGS_PHASE1.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 302558)
Muzny,D.M., Adams,C.C., Adio-Oduola,B., Ali-Isman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Binaige,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,

```

```

Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,M., Nickerson,E., Nwokenkwo,S.,
Ogih,M., Okuwonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,K., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 302558)
Worley,K.C.
Direct Submission
Submitted (22-APR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 17, 2000 this sequence version replaced gi:10047534.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HAYU
Center clone name: RP11-21B16
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 95% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 216542 bases at least Q40
Consensus quality: 251362 bases at least Q30
Consensus quality: 272372 bases at least Q20
Estimated insert size: 270959; sum-of-ctnigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.7x in Q20 bases; sum-of-ctnigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
consists of 75 ctnigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the ctnigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
* 1 10389: contig of 10389 bp in length
* 10390 10489: gap of unknown length
* 10490 24833: contig of 14344 bp in length
* 24834 24933: gap of unknown length
* 24934 33602: contig of 8669 bp in length
* 33603 33702: gap of unknown length
* 33703 43029: contig of 9327 bp in length
* 43030 43129: gap of unknown length
* 43130 53459: contig of 10330 bp in length
* 53460 53559: gap of unknown length

```

\* 53560 60735: contig of 1716 bp in length  
 \* 60836 60835: gap of unknown length  
 \* 69802: contig of 8967 bp in length  
 \* 69803 69802: gap of unknown length  
 \* 76183: contig of 6281 bp in length  
 \* 76184 76183: gap of unknown length  
 \* 76284 82151: contig of 5868 bp in length  
 \* 82152 82251: gap of unknown length  
 \* 82252 89697: contig of 7446 bp in length  
 \* 89698 89797: gap of unknown length  
 \* 89798 94730: contig of 4933 bp in length  
 \* 94731 94830: gap of unknown length  
 \* 94831 100596: contig of 5766 bp in length  
 \* 100597 100696: gap of unknown length  
 \* 100697 105812: contig of 5116 bp in length  
 \* 105813 105912: gap of unknown length  
 \* 105913 113093: contig of 7181 bp in length  
 \* 113094 113193: gap of unknown length  
 \* 113194 118164: contig of 4971 bp in length  
 \* 118165 118264: gap of unknown length  
 \* 118265 125193: contig of 6929 bp in length  
 \* 125194 125293: gap of unknown length  
 \* 125294 130564: contig of 5271 bp in length  
 \* 130565 130664: gap of unknown length  
 \* 130665 136299: contig of 5635 bp in length  
 \* 136300 136399: gap of unknown length  
 \* 136400 142500: contig of 6101 bp in length  
 \* 142501 142600: gap of unknown length  
 \* 142601 146623: contig of 4023 bp in length  
 \* 146624 146723: gap of unknown length  
 \* 146724 150634: contig of 3911 bp in length  
 \* 150635 150734: gap of unknown length  
 \* 150735 157354: contig of 6620 bp in length  
 \* 157355 157454: gap of unknown length  
 \* 157455 161824: contig of 4370 bp in length  
 \* 161825 161924: gap of unknown length  
 \* 161925 166911: contig of 4987 bp in length  
 \* 166912 167011: gap of unknown length  
 \* 167012 171639: contig of 4628 bp in length  
 \* 171640 171739: gap of unknown length  
 \* 171740 178016: contig of 6277 bp in length  
 \* 178017 178116: gap of unknown length  
 \* 178117 184085: contig of 5969 bp in length  
 \* 184086 184185: gap of unknown length  
 \* 184186 189227: contig of 5041 bp in length  
 \* 189228 189326: gap of unknown length  
 \* 189327 194366: contig of 5040 bp in length  
 \* 194367 194466: gap of unknown length  
 \* 194467 198530: contig of 4064 bp in length  
 \* 198531 198630: gap of unknown length  
 \* 198631 203086: contig of 4456 bp in length  
 \* 203087 203186: gap of unknown length  
 \* 203187 206934: contig of 3748 bp in length  
 \* 206935 207034: gap of unknown length  
 \* 207035 211278: contig of 4244 bp in length  
 \* 211279 211378: gap of unknown length  
 \* 211379 214720: contig of 3342 bp in length  
 \* 214721 214820: gap of unknown length  
 \* 214821 217723: contig of 2903 bp in length  
 \* 217724 217823: gap of unknown length  
 \* 217824 221500: contig of 3677 bp in length  
 \* 221501 221600: gap of unknown length  
 \* 221601 225644: contig of 4044 bp in length  
 \* 225645 225744: gap of unknown length  
 \* 225745 229281: contig of 3537 bp in length  
 \* 229282 229381: gap of unknown length  
 \* 229382 232302: contig of 2921 bp in length  
 \* 232303 232402: gap of unknown length  
 \* 232403 235101: contig of 2699 bp in length  
 \* 235102 235201: gap of unknown length  
 \* 235202 237598: contig of 2297 bp in length  
 \* 237599 240299: contig of 2701 bp in length

\* 240300 240399: gap of unknown length  
 \* 240400 242831: contig of 2432 bp in length  
 \* 242832 242931: gap of unknown length  
 \* 242932 246173: contig of 3242 bp in length  
 \* 246174 246273: gap of unknown length  
 \* 246274 248692: contig of 2419 bp in length  
 \* 248693 248792: gap of unknown length  
 \* 248793 251172: contig of 2380 bp in length  
 \* 251173 251272: gap of unknown length  
 \* 251273 254217: contig of 2945 bp in length  
 \* 254218 254317: gap of unknown length  
 \* 254318 256518: contig of 2201 bp in length  
 \* 256519 256618: gap of unknown length  
 \* 256619 257832: contig of 1214 bp in length  
 \* 257833 257932: gap of unknown length  
 \* 257933 260746: contig of 2814 bp in length  
 \* 260747 260846: gap of unknown length  
 \* 260847 262821: contig of 1975 bp in length  
 \* 262822 262921: gap of unknown length  
 \* 262922 264989: contig of 2068 bp in length  
 \* 264990 265089: gap of unknown length  
 \* 265090 266903: contig of 1814 bp in length  
 \* 266904 267003: gap of unknown length  
 \* 267004 268962: contig of 1959 bp in length  
 \* 268963 269062: gap of unknown length  
 \* 269063 270581: contig of 1519 bp in length  
 \* 270582 270681: gap of unknown length

Query Match 79.2%; Score 19.8; DB 72; Length 302558;

Best Local Similarity 91.3%; Pred. No. 1.3e+02; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gttgtggtatctgtgtttgaag 23

Db 192545 GTCGTGGAACGCTGTTGGAAG 192567

# RESULT 24

AC063930

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC063930 303717 bp DNA  
 Homo sapiens chromosome 3 clone RP11-514F2, \*\*\* SQUENCING IN  
 PROGRESS \*\*\*, 86 unordered pieces.

AC063930

AC063930.17 GI:11995500

HTG: HTGS\_PHASE1.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 303717)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,

Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,

Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,

Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,

Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,

Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,

Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,

Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,

Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,

Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,

Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,

Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,

Loulsegh,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,

Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,



Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M.,  
 Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,  
 Morgan, A., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,  
 Ogulu, M., Okunolu, G., Oragunye, N., Oviado, R., Pace, A., Payton, B.,  
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,  
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,  
 Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N.,  
 Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,  
 Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,  
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,  
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,  
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
 Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S.,  
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.  
 and Gibbs, R.

Direct Submission  
 Unpublished  
 2 (bases 1 to 303717)  
 Worley, K.C.

Submitted (22-APR-2000) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Dec 29, 2000 this sequence version replaced gi:11141926.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: HAYW  
 Center clone name: RP11-514F2  
 ----- Summary Statistics  
 Sequencing vector: M13; L08821  
 Chemistry: Dye-terminator Big Dye; 3% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 221091 bases at least Q40  
 Consensus quality: 252136 bases at least Q30  
 Consensus quality: 268230 bases at least Q20  
 Estimated insert size: 266010; sum-of-contigs estimation  
 Quality coverage: 0x in Q20 bases; agarose-fp estimation  
 Quality coverage: 2.6x in Q20 bases; sum-of-contigs estimation

## COMMENT

1 9894: contig of 9894 bp in length  
 9895 9894: gap of unknown length  
 9995 20133: contig of 10139 bp in length  
 20134 20233: gap of unknown length  
 20234 29489: contig of 9256 bp in length  
 29490 29589: gap of unknown length  
 29590 36626: contig of 7037 bp in length  
 36627 36726: gap of unknown length  
 36727 42239: contig of 5513 bp in length  
 42240 42339: gap of unknown length  
 42340 50165: contig of 7826 bp in length  
 50166 50265: gap of unknown length  
 50266 58725: contig of 8460 bp in length  
 58726 58825: gap of unknown length  
 58826 63134: contig of 6309 bp in length  
 63135 65234: gap of unknown length  
 65235 71642: contig of 6408 bp in length

71743 71742: gap of unknown length  
 71743 71753: contig of 5411 bp in length  
 71753 71753: gap of unknown length  
 71754 84462: contig of 7209 bp in length  
 84463 84562: gap of unknown length  
 84563 91495: contig of 6933 bp in length  
 91496 91595: gap of unknown length  
 91596 97457: contig of 5862 bp in length  
 97458 97557: gap of unknown length  
 97558 105030: contig of 7473 bp in length  
 105031 105130: gap of unknown length  
 110324 110324: contig of 5194 bp in length  
 110325 110424: gap of unknown length  
 110425 115824: contig of 5400 bp in length  
 115825 115924: gap of unknown length  
 121202 121302: contig of 5278 bp in length  
 121303 124865: contig of 3563 bp in length  
 124866 124966: gap of unknown length  
 124967 131022: contig of 6057 bp in length  
 131023 131122: gap of unknown length  
 131123 134925: contig of 3803 bp in length  
 134926 135025: gap of unknown length  
 135026 139917: contig of 4892 bp in length  
 139918 140017: gap of unknown length  
 140018 144220: contig of 4203 bp in length  
 144221 144320: gap of unknown length  
 144321 147439: contig of 3119 bp in length  
 147440 147539: gap of unknown length  
 147540 152505: contig of 4966 bp in length  
 152506 152605: gap of unknown length  
 152606 156007: contig of 3401 bp in length  
 156008 161617: contig of 5511 bp in length  
 161618 161717: gap of unknown length  
 161718 167424: contig of 5707 bp in length  
 167425 171814: contig of 4290 bp in length  
 171815 171914: gap of unknown length  
 171915 174719: contig of 2805 bp in length  
 174720 174819: gap of unknown length  
 174820 179373: contig of 4554 bp in length  
 179374 179473: gap of unknown length  
 179474 183064: contig of 3591 bp in length  
 183065 183164: gap of unknown length  
 183165 186280: contig of 3116 bp in length  
 186281 186380: gap of unknown length  
 186381 190385: contig of 4005 bp in length  
 190386 193709: contig of 3224 bp in length  
 193710 193809: gap of unknown length  
 193810 197580: contig of 3771 bp in length  
 197581 197680: gap of unknown length  
 197681 200912: contig of 3232 bp in length  
 200913 201012: gap of unknown length  
 201013 204830: contig of 3818 bp in length  
 204831 204930: gap of unknown length  
 204931 208131: contig of 3201 bp in length  
 208132 208231: gap of unknown length  
 208232 210922: contig of 2691 bp in length  
 210923 211022: gap of unknown length  
 211023 214347: contig of 3325 bp in length  
 214348 214447: gap of unknown length  
 214448 216949: contig of 2502 bp in length  
 216950 217049: gap of unknown length  
 217050 219757: contig of 2708 bp in length  
 219758 219857: gap of unknown length  
 219858 223103: contig of 3246 bp in length  
 223104 223203: gap of unknown length  
 223204 226009: contig of 2806 bp in length  
 226010 226109: gap of unknown length  
 226110 229109: contig of 3000 bp in length  
 229110 229209: gap of unknown length

```

* 229210 232471: contig of 3262 bp in length
* 232472 232571: gap of unknown length
* 232572 232584: contig of 2713 bp in length
* 232585 232588: gap of unknown length
* 232589 232653: contig of 3269 bp in length
* 232654 232753: gap of unknown length
* 232754 240129: contig of 1376 bp in length
* 240130 240229: gap of unknown length
* 240230 242092: contig of 1863 bp in length
* 242093 242192: gap of unknown length
* 242193 245672: contig of 3480 bp in length
* 245673 245772: gap of unknown length
* 245773 247971: contig of 2199 bp in length
* 247972 248071: gap of unknown length
* 248072 250703: contig of 2632 bp in length
* 250704 250803: gap of unknown length
* 250804 252676: contig of 1873 bp in length
* 252677 252776: gap of unknown length
* 252777 255037: contig of 2261 bp in length
* 255038 255137: gap of unknown length

Query Match
Best Local Similarity 79.2%; Score 19.8; DB 72; Length 303717;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gttgctggatcgtctgttgaag 23
Db 231708 GTTGCTGGACCTGCTTGAAG 231730

RESULT 25
AC014347/c
LOCUS      22329 bp      DNA      HTG      16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
           pieces.
ACCESSION  AC014347
VERSION     AC014347.1  GI:6436988
KEYWORDS   HTG; HTGS_PHASE2.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 22329)
            Adams,M. and Venter,J.C.
            Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
            This sequence was identified as CDM:10212385 by the submitter.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES   Location/Qualifiers
            source
              1..22329
              /organism="Drosophila melanogaster"
              /db_xref="taxon:7227"
BASE COUNT  6272 a 5051 c 4780 g 6226 t
ORIGIN

Query Match
Best Local Similarity 76.8%; Score 19.2; DB 63; Length 22329;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttatggtatcgtctgttgaagc 25
Db 1605 TTGCGGATGCTGCTTGAAGCG 1582

RESULT 26
AF025422

```

```

LOCUS      38000 bp      DNA      PRI      20-OCT-1997
DEFINITION Homo sapiens chromosome 19 cosmid F15386, genomic sequence,
           complete sequence.
ACCESSION  AF025422
VERSION     AF025422.1  GI:2547408
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 38000)
            Rank,D.R., Westphall,M.S., Ono,T., Berggren,W.T., Lamerdin,J.,
            Ashworth,L. and Smith,L.M.
            Direct Submission
            Submitted (15-SEP-1997) Department of Chemistry, University of
            Wisconsin-Madison, 1101 University Ave, Madison, WI 53706, USA
FEATURES   Location/Qualifiers
            source
              1..38000
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="F15386"
              /chromosome="19"
              /map="19q13.2"
              /cell_line="UV5HL9-5B"
              /clone_lib="LL19NC02 P2 chromosome 19-specific cosmid
              library"
              /note="cosmid library constructed at LLNL from flow-sorted
              chromosomes from hybrid UV5HL9-5B, which carries
              chromosome 19 as its only human chromosome."
            misc_feature
              273..435
              /note="Predicted exon, program: grill2exons_human_1.3,
              frame: 0, quality: good, score: 53.000"
              1322..1447
              /note="Predicted exon, program: grill2exons_human_1.3,
              frame: 1, quality: good, score: 60.000"
            misc_feature
              1456..1585
              /note="Predicted exon, program: grill2exons_human_1.3,
              frame: 1, quality: marginal, score: 48.000"
            misc_feature
              3163..3198
              /note="Predicted exon, program: grill2exons_human_1.3,
              frame: 0, quality: good, score: 55.000"
              3296..3495
              /note="Predicted exon, program: grill2exons_human_1.3,
              frame: 0, quality: excellent, score: 92.000"
            repeat_region
              4818..5102
              /rpt_family="L1"
            repeat_region
              5368..5528
              /rpt_family="MER21"
            repeat_region
              5629..5672
              /rpt_family="MER21"
            misc_feature
              complement(6262..6374)
              /note="Predicted exon, program: grill2exons_human_1.3,
              frame: 0, quality: good, score: 74.000"
            misc_feature
              complement(7274..7318)
              /note="Predicted exon, program: grill2exons_human_1.3,
              frame: 0, quality: excellent, score: 79.000"
            misc_feature
              complement(7733..7933)
              /note="Predicted exon, program: grill2exons_human_1.3,
              frame: 0, quality: good, score: 57.000"
            repeat_region
              complement(8100..8359)
              /rpt_family="ALU"
            repeat_region
              complement(8787..9087)
              /rpt_family="ALU"
            misc_feature
              complement(10868..11032)
              /note="Predicted exon, program: grill2exons_human_1.3,
              frame: 0, quality: marginal, score: 47.000"
            misc_feature
              complement(11524..11677)
              /note="Predicted exon, program: grill2exons_human_1.3,
              frame: 0, quality: good, score: 56.000"
            misc_feature
              complement(14364..14556)
              /note="Predicted exon, program: grill2exons_human_1.3,
              frame: 2, quality: good, score: 65.000"
            misc_feature
              complement(16428..16579)

```

```

/note="predicted exon, program: graill2exons_human_1.3,
frame: 0, quality: excellent, score: 76.000"
/rpt_family="ALU"
repeat_region
/rpt_family="ALU"
repeat_region
/rpt_family="ALU"
repeat_region
/rpt_family="MER21"
repeat_region
complement(21314..22104)
/rpt_family="L1"
repeat_region
complement(21990..22126)
/note="predicted exon, program: graill2exons_human_1.3,
frame: 2, quality: good, score: 52.000"
complement(22667..22933)
/rpt_family="ALU"
repeat_region
complement(23080..23300)
/rpt_family="ALU"
repeat_region
complement(24284..24629)
/rpt_family="THE1"
repeat_region
24753..24861
/note="predicted exon, program: graill2exons_human_1.3,
frame: 0, quality: good, score: 58.000"
25445..25663
/rpt_family="ALU"
repeat_region
complement(25695..26047)
/rpt_family="THE1"
repeat_region
26097..26382
/rpt_family="ALU"
repeat_region
complement(26593..26981)
/rpt_family="L1"
repeat_region
27076..27220
/note="predicted exon, program: graill2exons_human_1.3,
frame: 0, quality: excellent, score: 100.000"
complement(27316..27383)
/rpt_family="MLT1"
repeat_region
28274..28399
/note="predicted exon, program: graill2exons_human_1.3,
frame: 1, quality: good, score: 62.000"
28900..29086
/note="predicted exon, program: graill2exons_human_1.3,
frame: 2, quality: excellent, score: 88.000"
29251..31136
/note="BLASTX similarity to finger protein 2, placental -
human: ZINC FINGER PROTEIN ZFP-38 (C7FIN51) (TRANSCRIPTION
FACTOR RU49) >gnl|PIB|d1001955 (D10630) zinc finger
protein [Mus musculus]
EST similarities: T23166, T12758"
32217..32550
/rpt_family="THE1"
repeat_region
33424..33521
/note="predicted exon, program: graill2exons_human_1.3,
frame: 0, quality: excellent, score: 100.000"
complement(33811..33919)
/rpt_family="MLT1"
repeat_region
35479..38000
/rpt_family="L1"
repeat_region
35672..35800
/note="predicted exon, program: graill2exons_human_1.3,
frame: 1, quality: marginal, score: 46.000"
36135..36344
/note="predicted exon, program: graill2exons_human_1.3,
frame: 2, quality: good, score: 63.000"

```

BASE COUNT 10578 a 8454 c 7972 g 10996 t  
ORIGIN

Query Match 76.8%; Score 19.2; DB 88; Length 38000;  
Best Local Similarity 87.5%; Pred. No. 2.1e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gttgctggatctgctgtttgaagc 24

Db 11472 GTTGTGGCTGCTGCTTTTGTAGC 11495

RESULT 27

HSMX1\_1

WPCOMMENT

Sequence split into 7 fragments LOCUS HSMX1 Accession AJ011929

Fragment Name	Begin	End
HSMX1_0	1	110000
HSMX1_1	100001	210000
HSMX1_2	200001	310000
HSMX1_3	300001	410000
HSMX1_4	400001	510000
HSMX1_5	500001	610000
HSMX1_6	600001	613769

Continuation (2 of 7) of HSMX1 from base 100001 (AJ011929 Homo sapiens chromosome 21 s. 1/2000)

Query Match 76.8%; Score 19.2; DB 84; Length 110000;  
Best Local Similarity 87.5%; Pred. No. 2.3e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gttgctggatctgctgtttgaagc 24

Db 23621 GATGCTGTTCTGCTGTTTGAAGC 23644

RESULT 28

LOCUS

AC010111 138938 bp DNA HTG 18-FEB-2000  
Drosophila melanogaster clone RPC198-9B18, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 63 unordered pieces.

AC010111

VERSION AC010111.4 GI:6996750

KEYWORDS HTG; HTGS, PHASE1.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 138938)

AUTHORS

Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K.,  
Bodett, B., Bouck, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C.,  
Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,  
David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,  
Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,  
Forcum-Tansey, J., Frantz, P., Ganes, R., Gorrell, J.H., Gorrell, L.L.,  
Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hoques, M.,  
Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M.,  
Kelly, S., Kondejewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,  
Lichtarge, O., Liu, J., Liu, W., Logan, O., Lozano, R.J., Lu, J.,  
Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M.,  
Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S.,  
Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L.,  
Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,  
Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sugeng, R.,  
Taber, P., Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wahbah, M.,  
Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A.,  
Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and  
Gibbs, R.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 138938)

Worley, K.C.

TITLE Direct Submission

JOURNAL Submitted (13-SEP-1999)

Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Feb 18, 2000 this sequence version replaced gi:590227.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

```

Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: DRDY
Center clone name: RPL198-9b18
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 10% of reads
Assembly program: Phrap; version 0.980611
Consensus quality: 84003 bases at least Q40
Consensus quality: 104821 bases at least Q30
Consensus quality: 116061 bases at least Q20
Estimated insert size: 128604; sum-of-contigs estimation
Quality coverage: 1x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 63 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1114: contig of 1114 bp in length
* 1115: gap of unknown length
* 1135: contig of 907 bp in length
* 2041: contig of 2061 bp in length
* 2061: gap of unknown length
* 2062: contig of 1537 bp in length
* 3598: contig of 1537 bp in length
* 3599: gap of unknown length
* 3618: contig of 1380 bp in length
* 4998: contig of 1380 bp in length
* 5018: gap of unknown length
* 5019: contig of 971 bp in length
* 5989: contig of 971 bp in length
* 6009: gap of unknown length
* 7116: contig of 1107 bp in length
* 7136: gap of unknown length
* 8692: contig of 1556 bp in length
* 8712: gap of unknown length
* 8693: gap of unknown length
* 8713: contig of 1517 bp in length
* 10239: contig of 1517 bp in length
* 10240: gap of unknown length
* 11190: contig of 941 bp in length
* 11210: gap of unknown length
* 12151: contig of 941 bp in length
* 12171: gap of unknown length
* 12172: contig of 759 bp in length
* 12930: contig of 759 bp in length
* 12931: gap of unknown length
* 12950: gap of unknown length
* 14261: contig of 1311 bp in length
* 14262: gap of unknown length
* 14281: gap of unknown length
* 15628: contig of 1347 bp in length
* 15629: gap of unknown length
* 15648: gap of unknown length
* 16920: contig of 1272 bp in length
* 16921: gap of unknown length
* 16940: gap of unknown length
* 18079: contig of 1139 bp in length
* 18080: gap of unknown length
* 18099: gap of unknown length
* 18946: contig of 847 bp in length
* 18947: gap of unknown length
* 18966: gap of unknown length
* 19946: contig of 980 bp in length
* 19966: gap of unknown length
* 20787: contig of 821 bp in length
* 20788: gap of unknown length
* 20808: gap of unknown length
* 22768: contig of 1961 bp in length
* 22769: gap of unknown length
* 22788: gap of unknown length
* 23673: contig of 885 bp in length
* 23674: gap of unknown length
* 23693: gap of unknown length
* 23694: contig of 1539 bp in length
* 25232: contig of 1539 bp in length
* 25233: gap of unknown length
* 27310: contig of 2058 bp in length
* 27311: gap of unknown length
* 27330: gap of unknown length
* 28319: contig of 989 bp in length
* 28320: gap of unknown length
* 28340: contig of 1401 bp in length
* 29760: gap of unknown length
* 29761: contig of 1615 bp in length
* 31375: contig of 1615 bp in length
* 31376: gap of unknown length

```

```

31396
33359: contig of 1964 bp in length
33379: gap of unknown length
3380: contig of 1518 bp in length
34898
34917: gap of unknown length
34918
36591: contig of 1674 bp in length
36592
36611: gap of unknown length
36612
37630: contig of 1019 bp in length
37631
37650: gap of unknown length
37651
38842: contig of 1192 bp in length
38843
38862: gap of unknown length
38863
40413: contig of 1551 bp in length
40414
42428: contig of 1995 bp in length
42429
42448: gap of unknown length
42449
43991: contig of 1543 bp in length
43992
44011: gap of unknown length
44012
45215: contig of 1204 bp in length
45216
45235: gap of unknown length
45236
47171: contig of 1936 bp in length
47172
47191: gap of unknown length
47192
49075: contig of 1884 bp in length
49076
49095: gap of unknown length
49096
51980: contig of 2885 bp in length
51981
52000: gap of unknown length
52001
53551: contig of 1551 bp in length
53552
53571: gap of unknown length
56030: contig of 2459 bp in length
56031
56050: gap of unknown length
56051
57435: contig of 1385 bp in length
57436
57455: gap of unknown length
59333: contig of 1878 bp in length
59334
59353: gap of unknown length
61391: contig of 2038 bp in length
61392
61411: gap of unknown length
61412
64200: contig of 2789 bp in length
64201
64220: gap of unknown length
64221
66723: contig of 2503 bp in length
66724
66743: gap of unknown length
66744
68487: contig of 1744 bp in length
68488
68507: gap of unknown length
68508
70631: contig of 2124 bp in length
70632
70651: gap of unknown length
70652
73101: contig of 2450 bp in length
73102
73121: gap of unknown length
73122
76993: contig of 3872 bp in length
76994
77013: gap of unknown length
77014
80128: contig of 3115 bp in length
80129
80148: gap of unknown length
80149
81844: contig of 1696 bp in length
81845
81864: gap of unknown length
81865
83438
83457: gap of unknown length
83458
86257: contig of 2800 bp in length
86258
86277: gap of unknown length
86278
89119: contig of 2842 bp in length
89120
89139: gap of unknown length
89140
91732: contig of 2593 bp in length
91733
91752: gap of unknown length
91753
94902: contig of 3150 bp in length
94903
94922: gap of unknown length
94923
98773: contig of 3851 bp in length
98774
98793: gap of unknown length
98794
101186: contig of 2393 bp in length
101187
101206: gap of unknown length
101207
104287: contig of 3081 bp in length
104288
104307: gap of unknown length
104308
108276: contig of 3969 bp in length
108277
108296: gap of unknown length
108297
112998: contig of 4702 bp in length
113018: gap of unknown length
113019
120694: contig of 7676 bp in length
120695
120714: gap of unknown length
120715
127046: contig of 6332 bp in length

```

```

* 127047 127066: gap of unknown length
* 127067 138938: contig of 11872 bp in length.
FEATURES             Location/Qualifiers
     source            1..138938
     /organism="Drosophila melanogaster"
     /db_xref="taxon:7227"
     /clone="RPC198-9B18"
BASE COUNT  38349 a 29399 c 29584 g 40293 t 1313 others
ORIGIN
Query Match      76.8%; Score 19.2; DB 61; Length 138938;
Best Local Similarity 87.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY  2 ttctggatctgctgttgaagcg 25
      |||| |||| | |||| |||| ||||
Db 134576 TTCCGGATGCTCTGTTGAAGCG 134553

RESULT 29
AC084239/c
LOCUS      AC084239      176928 bp      DNA      PRI      18-OCT-2000
DEFINITION Homo sapiens chromosome 19, BAC CTC-512J12 (BC347040), complete
sequence.
ACCESSION  AC084239
VERSION    AC084239.1  GI:10864171
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 176928)
            Kodyianni,V., Ge,Y., Krummel,G.K., Kvikstad,E., Grable,L.,
            Severin,J., Gordon,L., Shannon,M., Brower,A., Olsen,A.S. and
            Smith,L.M.
            Sequence analysis of a 1mb region in 19q13.2 containing a zinc
            finger gene cluster
            Unpublished
REFERENCE  2 (bases 1 to 176928)
            Kodyianni,V., Ge,Y., Krummel,G.K., Kvikstad,E., Grable,L.,
            Severin,J., Gordon,L., Shannon,M., Brower,A., Olsen,A.S. and
            Smith,L.M.
            Direct Submission
            Submitted (18-OCT-2000) Department of Chemistry, University of
            Wisconsin, 1101 University Ave., Madison, WI 53706, USA
            Map and sequence oriented from centromere to q-telomere. BC347040
            (CTC-512J12) is adjacent to BC228680 (CTC-204F22, AC074331) on the
            left, and overlaps BC101503 (CTB-188D12, AC069278) on the right
            from bases 176,142 to 176,929 of this accession. Additional chr 19
            map and sequence information are available at:
            http://www.bio.lnl.lnln.gov/genome/genome.html.
FEATURES             Location/Qualifiers
     source            1..176928
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /chromosome="19"
     /map="19q13.2"
     /clone="CTC-512J12 (LLNL clone name BC347040)"
     /clone_lib="CIT-HSPC"
     complement(<1..>414)
     /gene="ZFP93"
     /product="ZFP93 mRNA, partial CDS and 3'UTR"
     complement(<1..>414)
     /gene="ZFP93"
     /note="zinc finger protein homologous to Zfp93 in mouse"
     complement(104..>412)
     /gene="ZFP93"
     /note="zinc finger protein homologous to Zfp93 in mouse;
     Complete human protein sequence not currently available.
     Location refers to currently available ZFP93 sequence"
     /codon_start=1
     /product="ZFP93"

```

```

/protein_id="AAG23969.1"
/db_xref="GI:10864173"
/translation="SQRSNLQVHQI IHTGEKPKCECGKFSNSAGLSAQRVHTGE
KPYTCQCGKGFSGQASHFTHQRVHTGERPYICDVCKGFSQSRSHLIYHQRVHTGNL
"
2778..2849
/rpt_family="L1M4"
complement(3326..3528)
/rpt_family="AluSg/x"
3830..3886
/rpt_family="A-rich"
3887..4262
/rpt_family="L1M4"
4348..4992
/rpt_family="L1M4"
4996..5138
/rpt_family="L1M4"
5133..6055
/rpt_family="L1PA15-16"
6080..8755
/rpt_family="L1"
complement(8808..8888)
/rpt_family="L1P3"
8895..9023
/rpt_family="L1MD3"
9559..9961
/rpt_family="L1M4"
complement(10360..10650)
/rpt_family="AluSp"
complement(10654..10723)
/rpt_family="L2"
10724..11008
/rpt_family="AluJb"
complement(11009..11277)
/rpt_family="L2"
complement(11278..11583)
/rpt_family="AluSx"
complement(11584..11636)
/rpt_family="L2"
12104..12186
/rpt_family="MER112"
12353..12465
/rpt_family="MER20"
13026..13244
/rpt_family="MIR"
complement(13281..13445)
/rpt_family="L1ME"
13446..13747
/rpt_family="AluSx"
complement(13748..13797)
/rpt_family="L1ME"
complement(14360..14650)
/rpt_family="AluSx"
15265..15288
/rpt_family="AT-rich"
15409..15632
/rpt_family="AluJo"
15633..16079
/rpt_family="(TA)n"
complement(16373..16532)
/rpt_family="MER91A"
16634..16768
/rpt_family="MIR"
complement(17464..17622)
/rpt_family="MIR"
complement(18321..18758)
/rpt_family="MER110A"
complement(18768..19594)
/rpt_family="LTR1"
19774..19927
/standard_name="NIB531"
/note="RH55859"
19776..19915
STS
STS

```

```

/standard_name="stsg40922"
/notes="RH78239"
complement(20076..20252)
/rpt_family="Trigger2a"
complement(20253..20489)
/rpt_family="AluSp"
complement(20490..20752)
/rpt_family="Trigger2a"
complement(20780..21037)
/rpt_family="L1MC5"
complement(21049..21454)
/rpt_family="LTR34"
complement(21455..21986)
/rpt_family="LTR8"
complement(21987..22292)
/rpt_family="AluYa5"
complement(22293..22451)
/rpt_family="LTR8"
complement(22452..22726)
/rpt_family="LTR34"
22938..23075
/rpt_family="MER77"
23099..23672
/rpt_family="MER77"
complement(23732..24027)
/rpt_family="AluJo"
complement(24056..24219)
/rpt_family="Trigger1"
24220..24506
/rpt_family="AluY"
complement(24507..25062)
/rpt_family="Trigger1"
25062..25121
/rpt_family="LTR16A"
25323..25386
/rpt_family="AT-rich"
complement(25703..25955)
/rpt_family="L1ME"
25956..26264
/rpt_family="AluSx"
complement(26265..26438)
/rpt_family="L1ME"
26863..26952
/rpt_family="L1M4"
26953..27441
/rpt_family="LTR42"
27442..27648
/rpt_family="L1M4"
28112..28304
/rpt_family="L1"
complement(28341..28631)
/rpt_family="AluSg"
complement(28864..28933)
/rpt_family="L1MD3"
complement(29473..31095)
/rpt_family="L1M4A"
31096..31127
/rpt_family="(TTAAA)n"
complement(31128..31138)
/rpt_family="L1M4A"
complement(31139..31428)
/rpt_family="AluY"
complement(31429..36250)
/rpt_family="L1M4A"
complement(36293..37911)
/rpt_family="L1MCC"
38687..38708
/rpt_family="AT-rich"
complement(join(39445..39934,40052..42824,49512..49607,
53304..53430,56142..56192))
/gene="ZNF228"
/product="ZNF228 mRNA"
complement(39445..56192)

```

---

```

/STS
Query Match 76.8%; Score 19.2; DB 88; Length 176928;
Best Local Similarity 87.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gttgctggatcgcgtgtttgaagc 24
||||| ||||| ||||| ||||| |||||
Db 160901 GTTGCTGGCTCTGCTTTTGTAGC 160878

RESULT 30
AC008224/c
LOCUS
DEFINITION
AC008224
AC008224.18 GI:12957655
HTG.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 181636)
AUTHORS
Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferrera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Hock,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Paclet,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Scheeler,F.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 3R, region 83D-83D
Unpublished
2 (bases 1 to 181636)
AUTHORS
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Paclet,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 17, 2001 this sequence version replaced gi:7239409.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this HAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgp@fruitfly.berkeley.edu.
Location/Qualifiers
1..181636
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="83D-83D"
FEATURES
source

```

```

/clone="BACR29J02 (D817)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACE3.6)"
BASE COUNT 51104 a 38903 c 38541 g 53088 t
ORIGIN

Query Match 76.8%; Score 19.2; DB 4; Length 181636;
Best Local Similarity 87.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgtgcatctgtgtttgaagc 25
||||| ||||| ||||| ||||| |||||
Db 148551 TTGCCGATGCTGCTTTGAAGCG 148528

RESULT 31
AC011822 184304 bp DNA HTG 04-SEP-2000
LOCUS Homo sapiens chromosome 11 clone RP11-2022 map 11, WORKING DRAFT
DEFINITION AC011822
ACCESSION AC011822
VERSION AC011822.5 GI:9966274
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 184304)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 11, clone RP11-2022
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 184304)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Horton,L.,
Galagan,J., Gardy,J., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (15-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 4, 2000 this sequence version replaced gi:8096937.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2543
Center clone name: 2_0_22
----- Summary Statistics
Sequencing vector: M13; M7815; 96% of reads
Sequencing vector: Plasmid; n/a; 0.0% of reads
3.9119804400978Cheministry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 179037 bases at least Q40
Consensus quality: 181715 bases at least Q30
Consensus quality: 182600 bases at least Q20
Insert size: 188000; agarose-fp
Insert size: 183604; sum-of-contigs

```

```

Quality coverage: 5.3 in Q20 bases; agarose-fp
Quality cover.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 17743: contig of 17743 bp in length
* 17744 17843: gap of 100 bp
* 17844 26355: contig of 8512 bp in length
* 26356 26455: gap of 100 bp
* 26456 39048: contig of 12593 bp in length
* 39049 39148: gap of 100 bp
* 39149 58914: contig of 19766 bp in length
* 58915 59014: gap of 100 bp
* 59015 100646: contig of 41632 bp in length
* 100647 100746: gap of 100 bp
* 100747 123068: contig of 22322 bp in length
* 123069 123168: gap of 100 bp
* 123169 164942: contig of 41774 bp in length
* 164943 165042: gap of 100 bp
* 165043 184304: contig of 19262 bp in length.

```

## FEATURES

Location/Qualifiers

1..184304

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="11"

/map="11"

/clone="RP11-2022"

/clone\_lib="RPCI-11 Human Male BAC"

1..17743

/note="assembly\_fragment"

clone\_end:SP6

vector\_side:left"

17844..26355

/note="assembly\_fragment"

26456..39048

/note="assembly\_fragment"

39149..58914

/note="assembly\_fragment"

59015..100646

/note="assembly\_fragment"

100747..123068

/note="assembly\_fragment"

123169..164942

/note="assembly\_fragment"

165043..184304

/note="assembly\_fragment"

clone\_end:T7

vector\_side:right"

vector\_end:T7

BASE COUNT 53699 a 37653 c 38843 g 53409 t 700 others

## ORIGIN

Query Match 76.8%; Score 19.2; DB 62; Length 184304;  
 Best Local Similarity 87.5%; Pred. No. 2.4e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gtgtgcatctgtgtttgaagc 24

||||| ||||| ||||| ||||| |||||

Db 66065 GTTGCTGGCTCTGCTTTTGTAGC 66088

## RESULT 32

## HSMX1A

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

HSMX1A 300050 bp DNA PRI 23-SEP-2000  
 Homo sapiens chromosome 21 from 5 PACs and 5 Cosmids map  
 21q22.2.D21S349-MX1; segment 1/2, complete sequence.  
 AL442166 AJ011929  
 AL442166.1 GI:10303259

```

KEYWORDS
SOURCE
ORGANISM
Homo sapiens

REFERENCE
AUTHORS
Ramser,J., Francis,F., Beck,A., Hennig,S., Klages,S., Borzym,K.,
Langer,I., Steffens,C., Hildmann,T., Dagand,E., Yaspo,M.,
Reinhardt,R. and Lehrach,H.

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Direct Submission
Submitted (01-OCT-1998) MPMG, Abt. Lehrach, Max Planck Institut
fuer Molekulare Genetik, Ihnestrasse 73, Berlin, 14195, Germany
Clones received from Resource Centre of the Human Genome Project at
the Max-Planck-Institut for Molecular Genetics.
COMMENT
Bases 300001..300050 overlap with AL442167.
FEATURES
source
1..300050
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.2.D21S349-WX1"
/clone="PAC RPCI-1 247E2"
/clone="PAC RPCI-1 146B4"
/clone="PAC RPCI-1 141D16"
/clone="PAC RPCI-1 269A14"
/clone="PAC RPCI-1 265B9"
/clone="cosmid LNLcl16 44C5"
/clone="cosmid LNLcl16 16H18"
/clone="cosmid LNLcl16 14C10"
/clone="cosmid LNLcl16 25D2"
/clone="cosmid LNLcl16 87D5"
/clone.lib="LL21NC02, Chromosome 21 specific cosmid
library, Lawrence Livermore National Laboratory (LLNL),
creator: Pieter de Jong; PAC: RPCI1,3-5, Roswell Park
Cancer Institute, creator: Pieter de Jong, P.Ioannou"

BASE COUNT
ORIGIN
88605 a 65311 c 64439 g 81695 t

Query Match 76.8%; Score 19.2; DB 93; Length 300050;
Best Local Similarity 87.58; Pred No. 2.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gttgctggatctgtgtttgaagc 24
||||| |||||||||

Db 122285 GATGCTGTTCTGCTGTTGAAGC 122308

RESULT 33
HS21C084
LOCUS
DEFINITION
Homo sapiens chromosome 21 segment HS21C084.
ACCESSION
AL163284 AP001739 BA000005
VERSION
AL163284.2 GI:71717380
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 340000)
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
Asakawa,S., Shintani,A., Sakaki,T., Nagamine,K., Mitsuyama,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstieck,G.,
Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
Reicheit,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,

```

```

TITLE
JOURNAL
Direct Submission
Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing
Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
Group * Institute of Molecular Biotechnology, Genome Analysis *
Keio University School of Medicine, Dept. of Molecular Biology *
GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
Genetics (addresses see below)
The Chromosome 21 Mapping and Sequencing Consortium consists of
* RIKEN Genomic Sciences Center, Human Genome Research Group, *
Sagamihara 228-8555, Japan,
* e.mail: sakaki@gs.riken.go.jp
* URL: http://hgp.gsc.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, *
Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Dept. of Molecular Biology, *
Tokyo 160-8582, Japan,
* e.mail: shimizu@db.med.keio.ac.jp
* URL: http://adenine.db.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
* Ihnestrasse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.fz-berlin.mpg.de/.
FEATURES
Source
1..340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
<1..48537
/Note="Accession No. AF043945"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P39C17, 5' partial"
/clone.lib="RPCI1,3-5 PAC library"
1803..2037
/Note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
complement(2442..2739)
/Note="AluSg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
2906..3301
/Note="L1M3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
3565..3599
/Note="(IG)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
3599..3618
/Note="(GA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
3712..3741
/Note="(CAAAA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM

```

```

Hennig,S., Riesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K., and
Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,K., and
Yaspo,M.L.
Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing
Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
Group * Institute of Molecular Biotechnology, Genome Analysis *
Keio University School of Medicine, Dept. of Molecular Biology *
GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
Genetics (addresses see below)
The Chromosome 21 Mapping and Sequencing Consortium consists of
* RIKEN Genomic Sciences Center, Human Genome Research Group, *
Sagamihara 228-8555, Japan,
* e.mail: sakaki@gs.riken.go.jp
* URL: http://hgp.gsc.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, *
Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Dept. of Molecular Biology, *
Tokyo 160-8582, Japan,
* e.mail: shimizu@db.med.keio.ac.jp
* URL: http://adenine.db.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
* Ihnestrasse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.fz-berlin.mpg.de/.
FEATURES
Source
1..340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
<1..48537
/Note="Accession No. AF043945"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P39C17, 5' partial"
/clone.lib="RPCI1,3-5 PAC library"
1803..2037
/Note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
complement(2442..2739)
/Note="AluSg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
2906..3301
/Note="L1M3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
3565..3599
/Note="(IG)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
3599..3618
/Note="(GA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
3712..3741
/Note="(CAAAA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM

```



```

repeat_region complement(4324..4371)
/nt="MLT2CB"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
repeat_region 4415..4446
/nt="(CA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region complement(4481..4917)
/nt="MLT2CB"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
repeat_region complement(5635..5747)
/nt="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
repeat_region 5972..6057
/nt="LINE"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
repeat_region complement(6421..6627)
/nt="MER3"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
repeat_region 8410..8615
/nt="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
repeat_region complement(9235..9341)
/nt="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
repeat_region complement(9523..9962)
/nt="MLT1C"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region complement(9995..10124)
/nt="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
repeat_region complement(10552..10739)
/nt="MER5A"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
repeat_region 10970..11104
/nt="HAL1"
/rpt_family="LINE/Other"
/rpt_type=DISPERSED
repeat_region 12600..13278
/nt="HAL1"
/rpt_family="LINE/Other"
/rpt_type=DISPERSED
repeat_region 13318..13662
/nt="HAL1"
/rpt_family="LINE/Other"
/rpt_type=DISPERSED
repeat_region 13663..13683
/nt="(TTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region complement(13684..13968)
/nt="AluX"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region 13969..14145
/nt="HAL1"
/rpt_family="LINE/Other"
/rpt_type=DISPERSED
repeat_region 14320..14340
/nt="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region 14387..14446

```

```

/nt="(TA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(15002..15087)
/nt="L1"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(15088..15392)
/nt="AluX"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(15393..16150)
/nt="L1"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(16150..17433)
/nt="L1PA16"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(17451..19052)
/nt="L1MB8"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(20290..20826)
/nt="MER1A"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
complement(20827..21658)
/nt="L1MB8"
/rpt_family="LINE/L1"

```

Query Match 76.8%; Score 19.2; DB 92; Length 340000;  
 Best Local Similarity 87.5%; Pred. No. 2.6e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gttgctggatctgctgtttgaagc 24

||||| ||||| ||||| ||||| |||||

Db 150650 GATGCTGTTCTGCTGTTTGAAGC 150673

```

RESULT 34
AF158246 550 bp DNA ROD 04-AUG-1999
LOCUS Cricetulus griseus glucose phosphate isomerase (GPI) gene, partial
DEFINITION Introns sequence.
ACCESSION AF158246
VERSION AF158246.1 GI:5690369
KEYWORDS Chinese hamster.
SOURCE Cricetulus griseus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Cricetulus
REFERENCE 1 (bases 1 to 550)
AUTHORS Williams,R.R.E., Hassan-Walker,A., Lavender,L., Morgan,M.M.J.,
Falk,P. and Ragoussis,I.
TITLE The minisatellite of the GPI/AMF/NLK/MF gene is conserved across
species and acts as an enhancer of an enhancer in vitro
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 550)
AUTHORS Williams,R.R.E., Hassan-Walker,A., Lavender,L., Morgan,M.M.J.,
Falk,P. and Ragoussis,I.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-1999), Medical and Molecular Genetics, King's
College London, Guy's Tower, Guy's Hospital, London SE1 9RT, UK
FEATURES
source Location/Qualifiers
db_xref="taxon:10029"

```

```

intron      <1..>550
gene         /gene="GPI"
            <1..>550
            /gene="GPI"
            /note="glucose phosphate isomerase"
BASE COUNT  110 a 126 c 176 g 138 t
ORIGIN

Query Match      75.2%; Score 18.8; DB 94; Length 550;
Best Local Similarity 90.9%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gttcgtggtctgctgtttgaa 22
    ||| ||||| ||||| |||||
Db 385 GTTCTGGATCTGCTGCTGAA 406

RESULT 35
AC021088      107024 bp  DNA      HTG      18-JUL-2000
LOCUS      Homo sapiens chromosome 5 clone CTD-2144A5, WORKING DRAFT SEQUENCE,
DEFINITION  8 ordered pieces.
ACCESSION  AC021088
VERSION    AC021088.3 GI:9256395
KEYWORDS  HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 107024)
AUTHORS   DOE Joint Genome Institute.
TITLES    Sequencing of Human Chromosome 5
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 107024)
AUTHORS   DOE Joint Genome Institute.
TITLES    Direct Submission
JOURNAL   Submitted (14-JAN-2000) Production Sequencing Facility, DOE Joint
           Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
           On Jul 18, 2000 this sequence version replaced gi:7711669.
COMMENT   -----Genome Center
           Center: Joint Genome Institute
           Center Code: JGI
           Web site: http://www.jgi.doe.gov
           -----
           Project Information
           Center Project Name: 682850
           Center clone name: CITB-HI_2144A5
           -----
           Summary Statistics
           Consensus quality: 100931 bases at least Q40
           Consensus quality: 105012 bases at least Q30
           Consensus quality: 105730 bases at least Q20
           Estimated insert size: 105000; pulse field gel estimation
           Quality coverage: 7.67 in Q20 bases; sum-of-contigs estimation
           Quality coverage: 7.55 in Q20 bases; sum-of-contigs estimation.
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 8 contigs. Gaps between the contigs
           * are represented as runs of N. The order of the pieces
           * is believed to be correct as given, however the sizes
           * of the gaps between them are based on estimates that have
           * been provided by the submitter.
           * This sequence will be replaced
           * by the finished sequence as soon as it is available and
           * the accession number will be preserved.
           *
           1 36057: contig of 36057 bp in length
           * 36058 36157: gap of unknown length
           * 36158 47839: contig of 11682 bp in length
           * 47840 47939: gap of unknown length
           * 47940 59311: contig of 11372 bp in length
           * 59312 59411: gap of unknown length
           * 59412 76365: contig of 16934 bp in length

```

---

```

* 76366 76465: gap of unknown length
* 76466 80920: contig of 4455 bp in length
* 80921 81020: gap of unknown length
* 81021 100116: contig of 19096 bp in length
* 100117 100216: gap of unknown length
* 100217 103745: contig of 3529 bp in length
* 103746 103845: gap of unknown length
* 103846 107024: contig of 3179 bp in length.
FEATURES             Location/Qualifiers
     source            1..107024
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2144A5"
                     /clone_lib="CalTech human BAC library D"
BASE COUNT  33920 a 21347 c 20863 g 30191 t 703 others
ORIGIN

Query Match      75.2%; Score 18.8; DB 66; Length 107024;
Best Local Similarity 90.9%; Pred. No. 3.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ttgctggtctgctgtttgaa 23
    ||||| ||||| ||||| |||||
Db 57296 TTGCTGAATCTGCTGTTGAAG 57317

RESULT 36
CNS06NLL/c
LOCUS      CNS06NLL      1029 bp  DNA      STS      10-JAN-2001
DEFINITION  T3 end of clone AU0AA015H11 of library AU0AA from strain CBS 3082
           of Saccharomyces kluyveri, sequence tagged site.
ACCESSION  AL406927
VERSION    AL406927.1 GI:12171587
KEYWORDS  Saccharomyces kluyveri.
SOURCE    Saccharomyces kluyveri.
ORGANISM  Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
           Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE  1 (bases 1 to 1029)
AUTHORS   Neuveglise, C., Bon, E., Lepingle, A., Wincker, P., Artiguenave, F.,
           Gaillardin, C., and Casaregola, S.
TITLES    Genomic Exploration of the Hemiascomycetous Yeasts: 9.
JOURNAL   Saccharomyces kluyveri
PUBMED   11152884
REFERENCE  2 (bases 1 to 1029)
AUTHORS   Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
           Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,
           de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,
           Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,
           Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
           Wincker, P., and Weissenbach, J.
TITLES    Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
           yeast species for molecular evolution studies(1)
JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
PUBMED   11152876
REFERENCE  3 (bases 1 to 1029)
AUTHORS   Genoscope.
TITLES    Direct Submission
JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
           2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
           seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
           This STS is part of a random genomic sequencing program of thirteen
           yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
           exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
           Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
           lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
           angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
           Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
           5 kb were prepared and both extremities were sequenced. See
           keywords for description of this sequence and for the sequence of

```

the other extremity of this insert.

```

FEATURES             source
    Location/Qualifiers
     1..1029
      /organism="Saccharomyces kluyveri"
      /strain="CBS 3082"
      /db_xref="taxon:4934"
      /clone="AU0AA015811"
      /clone_lib="AU0AA"
      /note="end : T3"
     1..1029
BASE COUNT      328 a 272 c 186 g 242 t 1 others
ORIGIN
    1..1029
      74.4%; Score 18.6; DB 53; Length 1029;
      Best Local Similarity 84.0%; Pred. No. 3.1e+02;
      Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gttgctgagctgctgtttgaagcg 25
      ||||| || ||||| |||||
Db 261 GTTGCTGCTCGCTGTTGTAGCG 237

RESULT 37
ATH249204/c          DNA          PLN          04-OCT-2000
LOCUS               4037 bp
DEFINITION          Arabidopsis thaliana pyk20 gene, promoter region.
ACCESSION           AJ249204
VERSION             AJ249204.1 GI:5824346
KEYWORDS            promoter; pyk20 gene.
SOURCE              thale cress.
ORGANISM            Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsids.
REFERENCE           1 (bases 1 to 4037)
AUTHORS             Puzio,P.S., Cai,D., Ohi,S., Wyss,U. and Grundler,F.M.W.
TITLE              Isolation of regulatory DNA regions related to differentiation of
nematode feeding structures in Arabidopsis thaliana
JOURNAL             Physiol. Mol. Plant Pathol. 53, 177-193 (1998)
AUTHORS             Puzio,P.S., Lausen,J., Heinen,P. and Grundler,F.M.
TITLE              Promoter analysis of pyk20, a gene from Arabidopsis thaliana
JOURNAL             Plant Sci. 157 (2), 245-255 (2000)
PUBMED              10960738
REFERENCE           3 (bases 1 to 4037)
AUTHORS             Puzio,P.S.
TITLE              Direct Submission
JOURNAL             Submitted (31-AUG-1999) Puzio P.S., Institut fuer Phytopathologie,
University Kiel, H. Rodewaldstr. 9, Kiel 24118, GERMANY
FEATURES             source
    Location/Qualifiers
     1..4037
      /organism="Arabidopsis thaliana"
      /strain="C-24"
      /db_xref="taxon:3702"
     1..4037
      /gene="pyk20"
      <1..4037
      /gene="pyk20"
      /citation=[1]
      /function="nematode responsive gene"
BASE COUNT      1286 a 638 c 658 g 1455 t
ORIGIN

Query Match      74.4%; Score 18.6; DB 13; Length 4037;
Best Local Similarity 84.0%; Pred. No. 3.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gttgctgagctgctgtttgaagcg 25
      ||||| ||||| ||||| |||||
Db 533 GTTGCTGAATCTGCTGTTGAGCG 509

```

```

RESULT 38
AC034199            DNA          PRI          29-SEP-2000
LOCUS               122877 bp
DEFINITION          Homo sapiens chromosome 5 clone CTFB-114C7, complete sequence.
ACCESSION           AC034199
VERSION             AC034199.5 GI:103337637
KEYWORDS            HTG.
SOURCE              human.
ORGANISM            Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE           1 (bases 1 to 122877)
AUTHORS             DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE              Direct Submission
JOURNAL             Unpublished
REFERENCE           2 (bases 1 to 122877)
AUTHORS             DOE Joint Genome Institute.
TITLE              Direct Submission
JOURNAL             Submitted (05-APR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE           3 (bases 1 to 122877)
AUTHORS             DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE              Direct Submission
JOURNAL             Submitted (29-SEP-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT            On Sep 29, 2000 this sequence version replaced gi:9256716.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.5% of Sequence;
Estimated Total Number of Errors is 0.4.
STS Content:
SHGC-104344 GS8254.
FEATURES             source
    Location/Qualifiers
     1..122877
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="5"
      /clone="CTFB-114C7"
BASE COUNT      33261 a 27391 c 27395 g 34830 t
ORIGIN

Query Match      74.4%; Score 18.6; DB 88; Length 122877;
Best Local Similarity 84.0%; Pred. No. 4.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gttgctgagctgctgtttgaagcg 25
      ||||| ||||| ||||| |||||
Db 57185 GTTGCTGAATCTGCTTTTACAGGG 57209

RESULT 39
AC016895/c          DNA          HTG          07-JUL-2000
LOCUS               168091 bp
DEFINITION          Homo sapiens chromosome 5 clone RP11-27P9, WORKING DRAFT SEQUENCE,
26 unordered pieces.
ACCESSION           AC016895
VERSION             AC016895.2 GI:7630807
KEYWORDS            HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE              human.
ORGANISM            Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE           1 (bases 1 to 168091)
AUTHORS             Waterston,R.H.
TITLE              The sequence of Homo sapiens clone
JOURNAL             Unpublished
REFERENCE           2 (bases 1 to 168091)
AUTHORS             Waterston,R.H.
TITLE              Direct Submission

```

## JOURNAL

Submitted (08-DEC-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Apr 21, 2000 this sequence version replaced gi:6539412.

## COMMENT

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0027P09
----- Summary Statistics -----
Sequencing vector: M13; 74%
Chemistry: Dye-primer ET; 26%
Sequencing vector: plasmid; 26%
Chemistry: Dye-terminator Big Dye; 26% of reads
Assembly program: Phrap; version 0.950319
Consensus quality: 153848 bases at least Q40
Consensus quality: 158311 bases at least Q30
Consensus quality: 160576 bases at least Q20
Insert size: 164000; agarose-fp
Insert size: 165591; sum-of-contigs
Quality coverage: 3.47 in Q20 bases; agarose-fp
Quality coverage: 3.49 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1063 1062: contig of 1062 bp in length
* 1163 1162: gap of unknown length
* 2458 2457: contig of 1295 bp in length
* 2558 2557: gap of unknown length
* 4705 4705: contig of 2148 bp in length
* 4805 4805: gap of unknown length
* 4806 6888: contig of 1883 bp in length
* 6889 6788: gap of unknown length
* 6789 8449: contig of 1661 bp in length
* 8450 8549: gap of unknown length
* 8550 9683: contig of 1134 bp in length
* 9684 9783: gap of unknown length
* 9784 11912: contig of 2129 bp in length
* 11913 12012: gap of unknown length
* 12013 13816: contig of 3804 bp in length
* 15817 15916: gap of unknown length
* 15917 18597: contig of 2681 bp in length
* 18598 18697: gap of unknown length
* 18698 22258: contig of 3561 bp in length
* 22259 22358: gap of unknown length
* 22359 24779: contig of 2421 bp in length
* 24780 24879: gap of unknown length
* 24880 27994: contig of 3115 bp in length
* 27995 28094: gap of unknown length
* 28095 33125: contig of 5031 bp in length
* 33126 33225: gap of unknown length
* 33226 37707: contig of 4482 bp in length
* 37708 37808: gap of unknown length
* 37808 43473: contig of 5666 bp in length
* 43474 43573: gap of unknown length
* 43574 49611: contig of 6038 bp in length
* 49612 49712: gap of unknown length
* 49712 58177: contig of 8465 bp in length
* 58177 58277: gap of unknown length
* 58277 65582: contig of 7306 bp in length
* 65583 65682: gap of unknown length
* 65683 73088: contig of 7406 bp in length
* 73089 73188: gap of unknown length
* 73189 82462: contig of 9274 bp in length
* 82463 82562: gap of unknown length

```

```

* 82563 91392: contig of 8830 bp in length
* 91393 91492: gap of unknown length
* 91493 103163: contig of 11671 bp in length
* 103164 103263: gap of unknown length
* 103264 115815: contig of 12552 bp in length
* 115816 115915: gap of unknown length
* 115916 131166: contig of 15251 bp in length
* 131167 131266: gap of unknown length
* 131267 147651: contig of 16385 bp in length
* 147652 147751: gap of unknown length
* 147752 168091: contig of 20340 bp in length.

FEATURES
      source
      1..168091
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="5"
        /clone="RP11-27P9"
      1..1062
        /note="assembly_name:Contig5"
      1163..2457
        /note="assembly_name:Contig6"
      2558..4705
        /note="assembly_name:Contig7"
      4806..6888
        /note="assembly_name:Contig8"
      6789..8449
        /note="assembly_name:Contig9"
      clone_end:T7
      vector_side:left
      8550..9683
        /note="assembly_name:Contig10"
      9784..11912
        /note="assembly_name:Contig11"
      12013..15816
        /note="assembly_name:Contig12"
      15917..18597
        /note="assembly_name:Contig13"
      18698..22258
        /note="assembly_name:Contig14"
      22359..24779
        /note="assembly_name:Contig15"
      24880..27994
        /note="assembly_name:Contig16"
      28095..33125
        /note="assembly_name:Contig17"
      33226..37707
        /note="assembly_name:Contig18"
      37808..43473
        /note="assembly_name:Contig19"
      43574..49611
        /note="assembly_name:Contig20"
      49712..58176
        /note="assembly_name:Contig21"
      58277..65582
        /note="assembly_name:Contig22"
      65683..73088
        /note="assembly_name:Contig23"
      73189..82462
        /note="assembly_name:Contig24"
      82563..91392
        /note="assembly_name:Contig25"
      91493..103163
        /note="assembly_name:Contig26"
      103264..115815
        /note="assembly_name:Contig27"
      115916..131166
        /note="assembly_name:Contig28"
      131267..147651
        /note="assembly_name:Contig29"
      clone_end:SP6
      vector_side:right
      147752..168091
        /note="assembly_name:Contig30"

```

BASE COUNT 51267 a 31894 c 31107 g 51315 t 2508 others  
ORIGIN

Query Match 73.6%; Score 18.4; DB 64; Length 168091;

Best Local Similarity 95.08; Pred. No. 5.8e+02;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ttgtcggtatcgtctgttga 21

|||||

Db 144862 TTGCTGGATCTGTTTGA 144843

RESULT 40

AF213884S1/c

LOCUS AF213884S1 190000 bp DNA PRI 21-FEB-2000  
DEFINITION Homo sapiens nuclear factor of kappa light polypeptide gene  
enhancer in B-cells 1 (NFKB1) gene, complete cds.

ACCESSION AF213884

VERSION AF213884.1 GI:7012904

KEYWORDS

SEGMENT

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

Chang, H.-M. and Tsai, S.-F.

TITLE

Hepatocellular Carcinoma Pathogenesis

Unpublished

2 (bases 1 to 190000)

Chang, H.-M. and Tsai, S.-F.

Direct Submission

Submitted (09-DEC-1999) Institute of Genetics, National Yang-Ming

University, 155 Li-Rong St. Section 2, Beitou, Taipei, Taiwan

11221, Republic of China

Location/Qualifiers

1. 190000

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="4"

/map="4q"

66142..66532

/gene="NFKB1"

/number=1

join(66142..66532,90256..90301,94580..94658,98592..98629,

102602..102700,131731..131879,141620..141783,

143625..143783,145279..145383,147610..147701,

149426..149564,158169..158312,159636..159725,

160882..161076,162264..162405,165639..165753,

171240..171441,171894..172063,172393..172495,

175319..175443,176788..176854,177178..177350,

178169..178325,181178..181659)

/gene="NFKB1"

/product="nuclear factor of kappa light polypeptide gene

enhancer in B-cells 1"

66142..181659

/gene="NFKB1"

90256..90301

/gene="NFKB1"

/number=2

join(90256..90301,94580..94658,98592..98629,

102602..102700,131731..131879,141620..141783,

143625..143783,145279..145383,147610..147701,

149426..149564,158169..158312,159636..159725,

160882..161076,162264..162405,165639..165753,

171240..171441,171894..172063,172393..172495,

175319..175443,176788..176854,177178..177350,

178169..178325,181178..181659)

/gene="NFKB1"

/note="p105; nuclear factor kappa-B DNA binding subunit"

/codon\_start=1

exon

mRNA

gene

exon

CDS

/product="nuclear factor of kappa light polypeptide gene  
enhancer in B-cells 1"  
/protein\_id="AAF35232.1"  
/db\_xref="GI:7012906"  
translation="MAEDDPYLGRPEQMFHLDPSLTHTIFNPEVFQPMALPTDGPYL  
OILEAPKQGRFRVYVCEGPHGLPGASSEKNKSYPOVKICNVVGPVAKVIVOLVTN  
CKNIHLHAHSLVGHKCEGICITVAGPKDMVGVGFANLGLHVTKKKVFTEARMTA  
CIRNGTGLLVHPDLAYLQAEGGGRQLGDRKELIRQAALQOKEMDLSSVRLMFTA  
FLPDSTGTRRLPEPVSDAIYDSKAPNASNIKIVMDORTACGVTGGEEIYLCQVQ  
KDDIQIRFEEENGWEGFGDFSPDVHROFAIVFKPKYKDIINITKPSVFEVQLR  
RKSLDTEPKPFLYPIKDEEVRQRKQKLPNFSDFGGSGAGAGGGGMEGSGG  
GGGTGSGPGVSPFHYGPTVGGITFHPTTKSNAGMKHGTMDTESKDGDCDS  
DKNTNLFKGVLETTEDQEPSEATVGNSEVLTATGTKEESAGVDNLFLXAMOL  
AKRHANALFYAVTGDVKMLLAVQRHLTAVQDENGDSVLHLAIHLHSQVRLLEVT  
SGLISDDIINRNDLYQTPHLAVITIKOEDVVVDLURAGADLSLLDRLGNVHLIAK  
EGHDLVLSILLKHKAAALLDHPNGDGLNAIHLAMNSLPCLLILLVAAGDAQEQ  
KSGRTALHLAVEHDNISLAGCLLLEGDAHVDSTTYDGTPLHIAAGRSTRIALLKA  
ACADPLVENFEPLDYLDPSWENAGEDGVPGTTPDMATSWQVDIILNGPYEFT  
SDLLAOGDMKOLAEADVKLQYKLLIIPDPKNWATLACKLGLGLNNAFLRSPAPSK  
TLMDNYESGGTVRELVEALROMGYTEALEVIOAASSPVKTTSOAHSIPLSPASTROQ  
IDELRDSVSDVSGVETSFRLKLSFTSLTSGASLLTLNKMHPHYCQEGPLEGKI"  
94580..94658  
/gene="NFKB1"  
/number=3  
98592..98629  
/gene="NFKB1"  
/number=4  
102602..102700  
/gene="NFKB1"  
/number=5  
131731..131879  
/gene="NFKB1"  
/number=6  
141620..141783  
/gene="NFKB1"  
/number=7  
143625..143783  
/gene="NFKB1"  
/number=8  
145279..145383  
/gene="NFKB1"  
/number=9  
147610..147701  
/gene="NFKB1"  
/number=10  
149426..149564  
/gene="NFKB1"  
/number=11  
158169..158312  
/gene="NFKB1"  
/number=12  
159636..159725  
/gene="NFKB1"  
/number=13  
160882..161076  
/gene="NFKB1"  
/number=14  
162264..162405  
/gene="NFKB1"  
/number=15  
165639..165753  
/gene="NFKB1"  
/number=16  
171240..171441  
/gene="NFKB1"  
/number=17  
171894..172063  
/gene="NFKB1"  
/number=18  
172393..172495  
/gene="NFKB1"  
/number=19  
175319..175443

```

/gene="NFKB1"
/number=20
176788. .176854
/gene="NFKB1"
/number=21
177178. .177350
/gene="NFKB1"
/number=22
178169. .178325
/gene="NFKB1"
/number=23
181178. .181659
/gene="NFKB1"
/number=24
189901. .190000
/note="sequence overlaps with 5' end of sequence in
GenBank Accession Number AF224669"
BASE COUNT 59885 a 37003 c 36056 g 57056 t
ORIGIN

Query Match      73.6%; Score 18.4; DB 89; Length 190000;
Best Local Similarity 95.0%; Pred. No. 5.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 tgcctgagctgctgtttgaa 22
|||||
Db 127950 TGCTGGATCTGCTTTTGAA 127931

RESULT 41
AC021120/c
LOCUS AC021120 192368 bp DNA HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 4 clone RP11-34708, WORKING DRAFT SEQUENCE,
15 unordered pieces.
ACCESSION AC021120
VERSION AC021120.3 GI:7231003
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 192368)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 192368)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Mar 13, 2000 this sequence version replaced gi:7107591.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: HNH0347008
----- Summary Statistics -----
Sequencing vector: M13; 69%
Chemistry: Dye-primer ET; 69% of reads
Chemistry: Dye-terminator Big Dye; 31% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 184232 bases at least Q40
Consensus quality: 186730 bases at least Q30
Consensus quality: 188009 bases at least Q20
Insert size: 206000; agarose-fp
Insert size: 190968; sum-of-contigs
Quality coverage: 4.35 in Q20 bases; agarose-fp
Quality coverage: 4.73 in Q20 bases; sum-of-contigs

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

* 1 1508: contig of 1508 bp in length
* 1509 1608: gap of unknown length
* 1609 3427: contig of 1819 bp in length
* 3428 3527: gap of unknown length
* 3528 7117: contig of 3590 bp in length
* 7118 7217: gap of unknown length
* 7218 12500: contig of 5283 bp in length
* 12501 12600: gap of unknown length
* 12601 17911: contig of 5311 bp in length
* 17912 18012: gap of unknown length
* 18013 24887: contig of 6876 bp in length
* 24888 24987: gap of unknown length
* 24988 31009: contig of 6022 bp in length
* 31010 31109: gap of unknown length
* 31110 39620: contig of 8511 bp in length
* 39621 39720: gap of unknown length
* 39721 47727: contig of 8007 bp in length
* 47728 47827: gap of unknown length
* 47828 57927: contig of 10100 bp in length
* 57928 58027: gap of unknown length
* 58028 68843: contig of 10816 bp in length
* 68844 68944: gap of unknown length
* 68945 80397: contig of 11453 bp in length
* 80398 80497: gap of unknown length
* 80498 92109: contig of 11512 bp in length
* 92110 92109: gap of unknown length
* 92110 130569: contig of 38461 bp in length
* 130570 130669: gap of unknown length
* 130670 192368: contig of 61699 bp in length.

```

## FEATURES

## source

```

1..192368
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-34708"
1..1508
/note="assembly_name:Contig7"
1609..3427
/note="assembly_name:Contig8"
3528..7117
/note="assembly_name:Contig9"
7218..12500
/note="assembly_name:Contig10"
12601..17911
/note="assembly_name:Contig11"
18012..24887
/note="assembly_name:Contig12"
24988..31009
/note="assembly_name:Contig13"
clone_end:SP6
vector_side:right
31110..39620
/note="assembly_name:Contig14"
39721..47727
/note="assembly_name:Contig15"
47828..57927
/note="assembly_name:Contig16"
58028..68843
/note="assembly_name:Contig17"
68944..80396
/note="assembly_name:Contig18"
80497..92008
/note="assembly_name:Contig19"
92109..130569

```

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

```

/note="assembly_name:Contig20"
130670..192368
/note="assembly_name:Contig21"
1420 others
BASE COUNT 59372 a 37302 c 35660 g 58614 t
ORIGIN

Query Match 73.6%; Score 18.4; DB 66; Length 192368;
Best Local Similarity 95.0%; Pred. No. 5.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ttgctggatctgctgtttgaa 22
|||||
Db 49316 TCGTGGATCTGCTTTTGAA 49297

RESULT 42
AE002676 2018 bp DNA INV 06-OCT-2000
LOCUS Drosophila melanogaster genomic scaffold 142000013385608, complete
DEFINITION
ACCESSION AE002676
VERSION AE002676.2 GI:10729318
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
REFERENCE
AUTHORS Adams M.D., Celisner, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,
Amanatides, P.G., Scher, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,
George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,
Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,
Brandon, R.C., Rogers, Y.H., Blazek, R.G., Champe, M., Pfeiffer, B.D.,
Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor
Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J., Basu, A.,
Andrews-Frannk, C., Baldwin, D., Ballew, R.M., Basu, A.,
Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y.,
Bencos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D.,
Botchan, M.R., Bouck, J.J., Brokstein, P., Brottier, P., Burtis, K.C.,
Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I.,
Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de
Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M.,
Dodson, K., Doup, L.E., Downes, N., Dugan-Rocha, S., Dunkov, B.C.,
Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferrier, S.,
Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S.,
Gelbart, W.M., Glasser, K., Glodek, A., Gong, P., Gorrell, J.H., Gu, Z.,
Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J.,
Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,
Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,
Kensil, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A.,
Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattel, B., McIntosh, T.C.,
McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C.,
Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,
Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,
Nusskern, D.R., Pacle, J.M., Palazzolo, M., Pittman, G.S., Pan, S.,
Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K.,
Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I.,
Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C.,
Stapleton, M., Strong, R., Sun, E., Svirska, R., Tector, C., Turner, R.,
Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A.,
Weinstock, G.M., Weissbach, J., Williams, S.M., Woodage, T.,
Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F.,
Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H.,
Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O.,
Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
REFERENCE 2 (bases 1 to 2018)
AUTHORS Adams, M.D., Celisner, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.

Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7289371.
Location/Qualifiers
1..2018
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 473 a 442 c 424 g 679 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 4; Length 2018;
Best Local Similarity 87.0%; Pred. No. 5.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ttgctggatctgctgtttgaagc 24
|||||
Db 1294 TTGCTGGATCTGCTGTTCGC 1316

RESULT 43
AC019735 2021 bp DNA HTG 03-JAN-2000
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
DEFINITION
ACCESSION AC019735
VERSION AC019735.1 GI:6665162
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
REFERENCE
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDL:10210774 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
Location/Qualifiers
1..2021
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 681 a 424 c 443 g 473 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 65; Length 2021;
Best Local Similarity 87.0%; Pred. No. 5.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ttgctggatctgctgtttgaagc 24
|||||
Db 728 TTGCTGGATCTGCTGTTCGC 706

RESULT 44
XLRNAHE XLRNAHE 2094 bp mRNA VPT 16-OCT-1992
LOCUS X.laevvis mRNA for an ATP dependent RNA helicase.
DEFINITION
ACCESSION X57328
VERSION X57328.1 GI:65059
KEYWORDS ATP-dependent RNA helicase.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
REFERENCE
AUTHORS Amphibia; Batrachia; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Mesobatrachia; Pipoidae; Pipidae;

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
FEATURES
source
1. .2094
/organism="Xenopus laevis"
/db_xref="taxon:8355"
1. .2094
/codon_start=1
/product="ATP dependent RNA helicase"
/protein_id="CAA40605.1"
/db_xref="GI:65060"
/db_xref="SWISS-PROT:P24346"
/translation="MSHVAVENVNLNLDQFAGLDLNSADAESGAGTKGRYIPPHLRN
KEASRNDNWDGSGGNGYINGQDDRDGRMGNDYDRGGYSGRTGDRGDFYDRENG
WNSGRDKDAYSFSGRSDRGKGSLSFNERGSGRRTDORQDFGDMGRSDKSGFGRF
DRGNSRWDNRDDEMDKPLAPNDVRVQELFSGSNTGINPEKVDIPVEATGSCNCP
HTESPVDYTMGEIIMGNLQTRYTPPTPVOKHAIPILILEKDLMACQTSKGTAAFL
LPILSOIYADQGPQAMKHLQNGRYGRKQFPLSILAPTRSLAVQIYEAKRATKRS
RVRCVYVGGADIGQIIRDLRGCHLLVATPGRVLDMMERKIGLDFCKYLVLDDEAR
MLDMGFEPQIRRIVEQDTPMPKQVQRTMMFSATPKEIQLARDFLDEYIFLAVGRVG
STENITQKVVMEEMDKRSFLDLNATGKDSLTLVFVETKKGADALEDFLYHEGYA
CTSHGDSORDEALHQRSGKSPILVATAVAARGLDISNVKHVINFDLPSPDIEY
VHRIGRTGRVGNGLATSFNFKNINIKDLILLVRAKQRPVPSLKNMAYEHOHKSS
SRGSRKSRFGGFGAKDYRQSSGSGFSGSRGSSGHGSGRFGGFGYGGFYNSDQY
GGNTGGSSQVDMGN"
BASE COUNT 605 a 383 c 568 g 538 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 8; Length 2094;
Best Local Similarity 87.0%; Pred. No. 5.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgctgagatctgctgtttgaagc 24
|||||
Db 1789 TTGCTGGATCTGCTGTGTGAAGC 1811

RESULT 45
AB029919/c
LOCUS
DEFINITION
AB029919 Mus musculus STAP mRNA for sperm tail associated protein, complete cds.
AB029919.1 GI:5525094
VERSION
AB029919.1
KEYWORDS
sperm tail associated protein; Sperm tail associated protein.
SOURCE
Mus musculus adult male testis cDNA to mRNA.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3429)
Ohuchi,J., Arai,T., Kon,Y. and Watanabe,T.
Sperm tail associated protein
Published Only in database (1999) In press
2 (bases 1 to 3429)
Ohuchi,J., Arai,T., Kon,Y. and Watanabe,T.
Direct Submission
TITLE
Submitted (12-JUL-1999) to the DDBJ/EMBL/GenBank databases.
JOURNAL
Tomomasa Watanabe, Laboratory of Experimental Animal Science,
Graduate School of Veterinary Medicine Hokkaido University; Kita 18
nishi 9 Kita ku, Sapporo, Hokkaido 060-0818, Japan
(E-mail:watanabe@vetmed.hokudai.ac.jp, Tel:81-11-706-5106(ex.5106),
Fax:81-11-717-7569)
FEATURES
Location/Qualifiers
1. .3429
/organism="Mus musculus"

```

```

/db_xref="taxon:10090"
/chromosome="8"
/dev_stage="adult"
/map="51-53"
/sex="male"
/tissue_type="testis"
/note="D8Mit198(51) - D8Mit138(53)"
92. .3160
/gene="STAP"
92. .3160
/gene="STAP"
/codon_start=1
/product="sperm tail associated protein"
/protein_id="BAA82514.1"
/db_xref="GI:5525095"
/translation="MLKLKGLRTAKGLKNEAGERDRDVSNLNLSKLLSLQDLTKNLHD
VCKRQKTIQENQLCVVEEAMKLANHNKQAQTLVFTDNQDMFRVNNKYHLRQIQQLKK
KLLTQOELEFRTQEQASVCSLQYQSLERQTSLLVLRHCKLKEDEVILYEPFM
GNHSGTGEKHLAQBLAGDKILSLERSNLRYRKYQTSLSLLECEOVKMLEG
ELSGLTQDPENKGDHPKRYIYTPCVIQPHOFTLKRSEVQKVSQEDDILQELRNK
LACSNLVLREPEALIKLOAEFASYTATHRHPPTSEDCEDITKILKHLQEKDSQCL
HVEEYNLVKDLMELEAVSEQKKIKMKMMKLELDLHGLREETSCTVIEKKQETVFL
QYRQDLQOQYTESQKLSLKKDLQDKDERLEHEKLMQVONSLEKEAELEKQ
TTKELDTQEARQSTKIDCEALRAEIQKLDSEAEQLEKLSQDNLTQSKEAHL
SASLEDARHKIENCILQKQKEEVIQKQSLQHLKQESKIEERKHNROLOELS
SELSEQRSLNAEKESLQKTLDEDEKKIDELPHSTOVSEKORELNSRKLREEE
LLEIKGLLEKREFOLKSEOKALEEFTALROAKRKKAKHLRKLREKENIQ
AELTSSSHLSDSLANKYNSOKVIOELNAEIAKQDSIMLQTLDSAIQKKNCFON
MYSKEAYEELVRKSGNQDQLTQALEKLTQATSEKLSNRQQTQERKAQLEDEIAA
YERMKKLNWELKIQGFQOQSEQEVHNEFKKLEEMSNQVLOWRQHQSDLMKAAKE
IQAREPEEMATLRESLADKEPEYPPAKLTPEKTRHRENDQIMNNVEQWAKAEQK
SOLNEKLGNLREQVYIAKLTQEKDHLHNMVHLQENKKLKNEIEKKLKAGNPRIC
VKAFPPNKLEPSQKGLCCALGWRGIPQDLTPKTDHKYMGLPHSSGSSVC"
380. .508
/gene="STAP"
/note="LEUCINE ZIPPER MOTIF"
1341. .1426
/gene="STAP"
/note="LEUCINE ZIPPER MOTIF"
2891. .2956
/gene="STAP"
/note="LEUCINE ZIPPER MOTIF"
3369. .3374
polyA_signal
polyA_site
BASE COUNT 1118 a 805 c 920 g 586 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 94; Length 3429;
Best Local Similarity 87.0%; Pred. No. 5.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

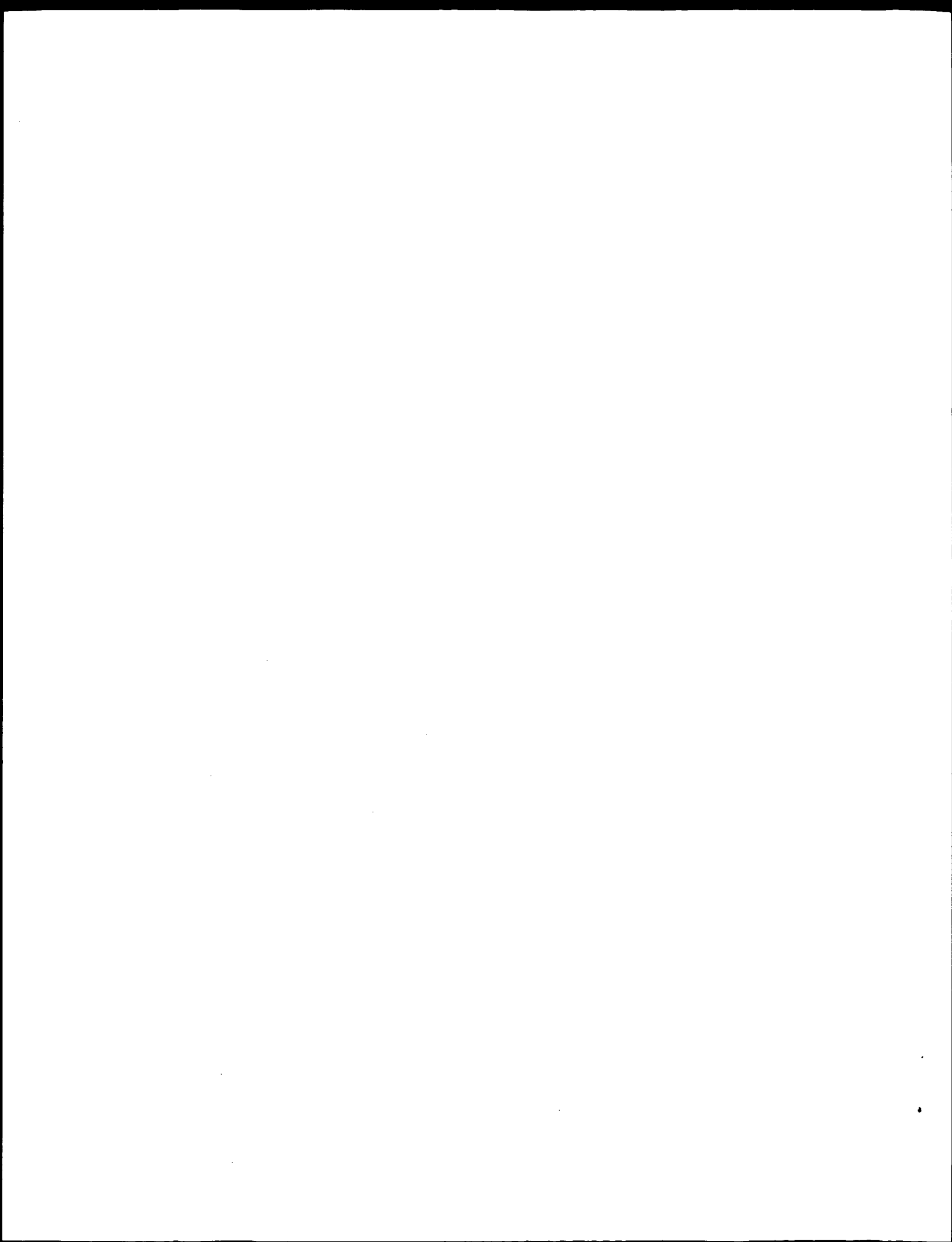
QY 2 ttgctgagatctgctgtttgaagc 24
|||||
Db 2557 TTGCTGCTGCTGCTGTGAAGC 2535

Search completed: October 9, 2001, 12:10:31
Job time: 3746 sec

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:36:16 ; Search time 470.56 Seconds  
(without alignments)  
33.359 Million cell updates/sec

Title: US-09-396-196F-3

Perfect score: 25

Sequence: 1 gttgctgacgtctgtttgaagcg 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_0601.\*  
1: /SIDSL/gcgdata/geneseq/geneseqn/NA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseqn/NA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseqn/NA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseqn/NA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseqn/NA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseqn/NA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseqn/NA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseqn/NA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseqn/NA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseqn/NA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseqn/NA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseqn/NA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseqn/NA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseqn/NA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseqn/NA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseqn/NA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseqn/NA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseqn/NA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseqn/NA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseqn/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match %	Length	ID	Description
1	25	100.0	1041	20	AXX01303
2	25	100.0	1084	10	AA91329
3	25	100.0	1121	7	AA60496
4	25	100.0	5872	15	AAQ62386
5	17.6	70.4	456	18	AA49104
6	17.6	70.4	527	21	AA55976
7	17.6	70.4	774	21	AAZ53871
8	17.6	70.4	1359	22	AA72295
9	17.6	70.4	3043	21	AA78167
10	17.6	70.4	4573	20	AA33946
11	17.6	70.4	4610	22	AA59633

c	12	17.6	70.4	52253	21	AA881478	N. meningitidis pa
	13	17.6	70.4	349880	21	AA21544	Neisseria meningit
	14	17.6	70.4	1437668	21	AA81490	N. meningitidis B
c	15	17.4	69.6	1994	21	AA21149	Human low adenosin
	16	17.4	69.6	1994	21	AA81426	Human I-kappa-B ki
c	17	17.4	69.6	1994	21	AAA35027	Human adenosine re
	18	17.4	69.6	2009	20	AA35027	Human RIP-associat
c	19	17.4	69.6	2034	20	AA207513	Human RIP-associat
	20	17.4	69.6	8631	21	AA207514	Human low adenosin
c	21	17.4	69.6	8631	21	AA35028	Human adenosine re
	22	17.2	68.8	434	21	AA35028	Human secreted pro
c	23	17.2	68.8	687	21	AA13715	Aspergillus oryzae
	24	17.2	68.8	1068	21	AA295242	Cyclophilin-type p
c	25	17.2	68.8	10564	16	AA203854	Human fibrinogen g
	26	17.2	68.8	14822	20	AA20543	Polynucleotide seq
c	27	17.2	68.8	1230025	20	AA91990	Nucleotide sequenc
	28	17	68.0	157	21	AA12134	Human secreted pro
	29	17	68.0	272	21	AA87515	Rat hepatocyte car
	30	17	68.0	492	21	AA44161	Arabidopsis thalia
	31	17	68.0	729	20	AA87783	EST clone ES306.
	32	17	68.0	751	21	AA58914	DNA encoding ketos
	33	17	68.0	925	20	AAV88879	EST clone H2103.
c	34	17	68.0	1008	18	AA770131	Max-interacting pr
	35	17	68.0	1200	21	AA48910	Arabidopsis thalia
	36	17	68.0	1237	21	AA34154	Arabidopsis thalia
	37	17	68.0	1514	20	AA39643	Renal cancer assoc
	38	17	68.0	1573	21	AA48922	Arabidopsis thalia
	39	17	68.0	1574	21	AA37600	Arabidopsis thalia
	40	17	68.0	2750	22	AA88117	Human FLEXHT-48 nu
	41	17	68.0	3435	17	AA735869	Human DNA polymera
c	42	17	68.0	4072	21	AA51054	Arabidopsis thalia
	43	16.8	67.2	309	20	AAV86759	EST clone AX318.
	44	16.8	67.2	1328	19	AAV58754	Human secreted pro
	45	16.8	67.2	12284	11	AAQ06001	Sequence encoding

#### ALIGNMENTS

RESULT 1  
AXX01303  
ID AXX01303 standard; DNA; 1041 BP.  
XX AXX01303;  
XX AXX01303;  
DT 12-APR-1999 (first entry)  
XX  
DE E. coli biotin synthetase (BioB) coding sequence.  
XX  
KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;  
KW biotin synthase; biotin production; vitamin H; BioB; ss.  
XX Escherichia coli.  
OS  
PN US5869719-A.  
XX  
PD 09-FEB-1999.  
XX  
PF 30-APR-1997; 97US-0846338.  
XX  
PR 30-APR-1997; 97US-0846338.  
PR 08-MAR-1995; 95US-0401068.  
XX (NOVS) NOVARTIS FINANCE CORP.  
PA Patton DA;  
PI  
XX WPI; 1999-152902/13.  
DR P-PSDB; AAW73906.  
DR  
PT Transgenic plants with high biotin levels - transformed with DNA  
PT encoding di:amino-pelargonic acid amino-transferase or biotin  
PT synthase

XX Example 2; Column 37-40; 34pp; English.

CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can  
 CC be used in the transgenic plant of the invention. The transgenic plant,  
 CC plant cell or plant tissue is transformed with a chimeric gene encoding  
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and  
 CC produces more biotin than a non-transgenic plant, cell or tissue. The  
 CC plant is used as an improved dietary source of biotin (vitamin H) for  
 CC humans or animals.

XX Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

SQ

Query Match 100.0%; Score 25; DB 20; Length 1041;  
 Best Local Similarity 100.0%; Pred. No. 0.11; Mismatches 0; Indels 0; Gaps 0;  
 Matches 25; Conservative 0;

QY 1 gttgctggtctgctgtttgaagcg 25  
 |||||

Db 57 gttgctggtctgctgtttgaagcg 81  
 |||||

RESULT 2

AAAN91329  
 ID AAAN91329 standard; DNA; 1084 BP.

XX AC AAAN91329;

XX DT 15-FEB-1990 (first entry)

XX DE E.coli Bio B gene.

XX KW E.coli; Bio B gene; biotin.

XX OS Escherichia coli.

XX FH Key Location/Qualifiers

XX FT CDS 24..1064

XX FT /\*tag=a

XX PN GB2216530-A.

XX PD 11-OCT-1989.

XX PF 17-MAR-1989; 89GB-0006210.

XX PR 22-MAR-1988; 88GB-0006804.

XX PR 17-MAR-1989; 89GB-0006210.

XX PA (UKAG-) UK MIN. AGRIC. FISH.

XX PI Pearson BM, McKee RA;

XX PS WPI; 1989-295085/41. P-PSDB P91392

DR Plasmid contg. gene(s) for expression of biotin synthetase enzymes  
 PT - derived from E.coli and capable of replication and expression in other  
 PT microorganisms, esp. yeast.

XX Table 3; page 33-4; 52pp; English.

XX The gene can be used in a plasmid for expression of enzymes of the biotin  
 CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae  
 CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for  
 CC Lactobacillus. Insertion of bio B improves biotin yields in  
 CC microorganisms which export biotin, or enables growth in media contg.  
 CC little or no biotin of organisms unable to synthesise biotin for their  
 CC own use.

XX Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

SQ

Query Match 100.0%; Score 25; DB 10; Length 1084;  
 Best Local Similarity 100.0%; Pred. No. 0.11; Mismatches 0; Indels 0; Gaps 0;  
 Matches 25; Conservative 0;

QY 1 gttgctggtctgctgtttgaagcg 25  
 |||||

Db 80 gttgctggtctgctgtttgaagcg 104  
 |||||

RESULT 3

AAAN60496  
 ID AAAN60496 standard; DNA; 1121 BP.

XX AC AAAN60496;

XX DT 17-OCT-1991 (first entry)

XX DE Sequence encoding biotin synthesising enzyme.

XX KW Biotin synthetic enzyme; E.coli; desthiobiotin; ds.

XX FH Key Location/Qualifiers

XX FT CDS 42..1082

XX FT /\*tag=a

XX PN JP61149091-A.

XX PD 07-JUL-1986.

XX PF 24-DEC-1984; 84JP-0272605.

XX PR 24-DEC-1984; 84JP-0272605.

XX PA (NIPS) NIPPON SODA KK.

XX DR WPI; 1986-216622/33.

XX DR P-PSDB; AAP60536.

XX PT Double stranded DNA encoding biotin synthesising enzyme -  
 PT comprises transformed mutant E.coli strain contg. cyclic doubled  
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.

XX PS Disclosure; Page 534; 23pp; Japanese.

XX CC The sequence may be expressed by a transformed E.coli host, cultured  
 CC in a medium containing desthiobiotin.

XX SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;  
 Best Local Similarity 100.0%; Pred. No. 0.11; Mismatches 0; Indels 0; Gaps 0;  
 Matches 25; Conservative 0;

QY 1 gttgctggtctgctgtttgaagcg 25  
 |||||

Db 98 gttgctggtctgctgtttgaagcg 122  
 |||||

RESULT 4

AAQ62386  
 ID AAQ62386 standard; DNA; 5872 BP.

XX AC AAQ62386;

XX DT 16-NOV-1994 (first entry)

XX DE Biotin-biosynthesis genes contg. plasmid pB030A-15/9.

XX KW Biotin; expression; enterobacteria; vitamin H; synthesis;  
 KW plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;  
 KW promoter ptac; biotin synthase; KAPA synthase;  
 KW 8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;

KW dethiobiotin synthase: DAPA synthase;  
 KW S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;  
 XX seborrhea; dermatitis; ds.  
 OS Escherichia coli DSM498.  
 FH Key Location/Qualifiers  
 FT Promoter 1..96  
 FT /tag= a  
 FT /function= "promoter ptac"  
 FT /evidence= EXPERIMENTAL  
 FT 23..28  
 FT /tag= b  
 FT /standard\_name= "promoter ptac"  
 FT 45..50  
 FT /tag= c  
 FT /evidence= EXPERIMENTAL  
 FT /standard\_name= "promoter ptac"  
 FT 105..109  
 FT /tag= d  
 FT /evidence= EXPERIMENTAL  
 FT /standard\_name= "bioB RBS no. 9"  
 FT 117..1157  
 FT /tag= e  
 FT /product= "biotin synthase"  
 FT /evidence= EXPERIMENTAL  
 FT /gene= "bioB"  
 FT /number= 1  
 FT 1141..1146  
 FT /tag= f  
 FT /standard\_name= "bioF RBS"  
 FT 1154..2311  
 FT /tag= g  
 FT /EC\_number= 2.3.1.47  
 FT /product= "KAPA synthase"  
 FT /evidence= EXPERIMENTAL  
 FT /gene= "bioF"  
 FT /number= 2  
 FT /standard\_name= "8-amino-7-oxononanoate synthase"  
 FT 2284..2288  
 FT /tag= h  
 FT /standard\_name= "bioC RBS"  
 FT 2295..3050  
 FT /tag= i  
 FT /function= "involved in pimeloyl-CoA synthesis"  
 FT /product= "protein"  
 FT /gene= "bioC"  
 FT /number= 3  
 FT 3030..3033  
 FT /tag= j  
 FT /standard\_name= "bioD RBS"  
 FT 3043..3753  
 FT /tag= k  
 FT /EC\_number= 6.3.3.3  
 FT /product= "DTB synthase"  
 FT /evidence= EXPERIMENTAL  
 FT /gene= "bioD15"  
 FT /number= 4  
 FT /standard\_name= "dethiobiotin synthase"  
 FT 3712..3750  
 FT /tag= l  
 FT /note= "bioD15 substitution"  
 FT 3742..3746  
 FT /tag= m  
 FT /standard\_name= "bioA RBS"  
 FT 3750..5039  
 FT /tag= n  
 FT /EC\_number= 2.6.1.62  
 FT /product= "DAPA synthase"  
 FT /evidence= EXPERIMENTAL  
 FT /gene= "bioA"  
 FT /number= 5  
 FT /standard\_name= "S-adenosyl-L-methionine: 8-amino-

FT RBS 5088..5093  
 FT /tag= o  
 FT /standard\_name= "ORFI RBS"  
 FT 5098..5574  
 FT /tag= p  
 FT /function= "unknown, involved in biotin synthesis"  
 FT /product= "protein"  
 FT /evidence= EXPERIMENTAL  
 FT /gene= "ORFI"  
 FT /number= 6  
 FT 5583..5644  
 FT /tag= q  
 FT /standard\_name= "rho-independent transcriptional  
 FT terminator"  
 FT 5583..5605  
 FT /tag= r  
 FT WO9408023-A.  
 FT 14-APR-1994.  
 FT 01-OCT-1993; 93WO-EP02688.  
 FT 02-OCT-1992; 92CH-0003124.  
 FT 15-JUL-1993; 93CH-0002134.  
 FT (LONZ ) LONZA AG.  
 FT Birch O, Brass J, Fuhrmann M, Shaw N;  
 FT WPI: 1994-135587/16.  
 FT P-PSDB; AARS1883, AARS1884, AARS1885, AARS1886, AARS1887, AARS1888.  
 FT Biotechnological biotin prodn. using enterobacterial biotin-gene  
 FT - providing vitamin H in high yield  
 FT Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.  
 FT The sequence is derived from plasmid pB030A-15/9 contg. the  
 FT bioB, bioF, bioC, and bioA genes responsible for biosynthesis  
 FT of biotin, arranged in a transcription unit. Microorganisms  
 FT contg. these DNA fragments or plasmids may be used in the prodn.  
 FT of biotin. Biotin (Vitamin H) may prevent seborrhea, dermatitis,  
 FT loss of appetite and tiredness.  
 FT SQ Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;  
 Query Match 100.0%; Score 25; DB 15; Length 5872;  
 Best Local Similarity 100.0%; Pred. No. 0.13; 0; Indels 0; Gaps 0;  
 Matches 25; Conservative 0; Mismatches 0;  
 Qy 1 gttgctgactgctgtgtttgaagcg 25  
 Db 173 gttgctgactgctgtgtttgaagcg 197  
 RESULT 5  
 ID AAT49104  
 XX AAT49104 standard; DNA; 456 BP.  
 AC AAT49104;  
 XX 23-SEP-1997 (first entry)  
 DE Partial DNA clone Acii#1-426 encoding immunostimulatory peptide.  
 KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;  
 KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;  
 KW tuberculin skin test; ds.  
 XX Mycobacterium tuberculosis.  
 OS

XX PN W09700067-A1.  
 XX PD 03-JAN-1997.  
 XX PF 14-JUN-1996; 96WO-US10375.  
 XX PR 15-JUN-1995; 95US-0000254.  
 XX PA (UYVI-) UNIV VICTORIA.  
 XX PI Nano FE;  
 XX DR WPI; 1997-077347/07.  
 XX PT New immuno:stimulatory peptide(s) of Mycobacterium tuberculosis -  
 XX PT useful in vaccines, diagnostic skin test, immunoassay and gene  
 XX PT isolation  
 XX PS Claim 1; Page 34; 79pp; English.  
 XX AA49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,  
 CC which encode partial sequences of immunostimulatory peptides. Each of  
 CC the clones encode at least one immunostimulatory T cell epitope. The  
 CC clones were identified by testing over 300 fusion clones (alkaline  
 CC phosphatase-M. tuberculosis peptide fusions) for their ability to  
 CC stimulate interferon (IFN)-gamma production. 80 clones were initially  
 CC designated to have some ability to stimulate IFN-gamma production, of  
 CC which 76 are shown in AA49100-175. These sequences can be used to  
 CC obtain the full length M. tuberculosis genes and corresponding proteins  
 CC using standard techniques. The peptides are useful in vaccines, as  
 CC reagents in an improved tuberculin skin test (especially using peptides  
 CC different from those used in vaccines so as to allow differentiation  
 CC between vaccinated and infected subjects) and as immunoassay reagents  
 CC for detecting specific antibodies. An advantage of these peptides is  
 CC that they stimulate production of IFN-gamma (critical for a protective  
 CC immune response to M. tuberculosis) by CD4-positive T cells. The protein  
 CC encoded by this sequence has amino acid similarity to a dipeptide  
 CC transport protein.  
 XX SQ Sequence 456 BP; 63 A; 121 C; 165 G; 98 T; 9 other;  
 Query Match 70.4%; Score 17.6; DB 18; Length 456;  
 Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 gttgctggtatctgctgtttgaagc 24  
 ||| ||||| ||||| ||| ||  
 Db 280 gttgctggtatctgctgtatgcgc 303  
 RESULT 6  
 AAC55976/C  
 ID AAC55976 standard; DNA; 527 BP.  
 XX AC AAC55976;  
 XX DT 25-JAN-2001 (first entry)  
 XX DE Eucalyptus grandis transcription factor DNA sequence #107.  
 XX KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;  
 KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;  
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;  
 KW homeodomain zipper; LIM domain; AP2; EREBS; zinc finger domain;  
 KW type 2 Cys2His2; CCAAT box element; MYB; ss.  
 XX OS Eucalyptus grandis.  
 XX PN W0200053724-A2.  
 XX PD 14-SEP-2000.

XX 09-MAR-2000; 2000WO-US06112.  
 XX 11-MAR-1999; 99US-0266513.  
 XX 18-AUG-1999; 99US-0149485.  
 XX (GENE-) GENESIS RES & DEV CORP LTD.  
 XX PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
 XX PI Wood M, McGrath A, Shenk MA, Glenn M;  
 XX WPI; 2000-579369/54.  
 XX PT New isolated polynucleotide encoding a plant transcription factor for  
 XX PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
 XX PT having modified gene expression or modified activity of a polypeptide  
 XX PT -  
 XX PS Claim 1; Page 72; 747pp; English.  
 XX CC The present invention relates to novel plant transcription factors from  
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding  
 CC sequence for one such transcription factor. The transcription factor may  
 CC be used to produce a plant having modified gene expression such as a  
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or  
 CC mahogany species or to modify the activity of a polypeptide in a plant.  
 CC The transcription factors of the present invention are members from the  
 CC following families of regulatory proteins: bZIP, bZIP family of G-box  
 CC binding factors, basic helix-loop-helix zipper,  
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2  
 CC and EREBS, zinc finger domains of type 2 Cys2His2, CCAAT box elements  
 CC and MYB.  
 XX SQ Sequence 527 BP; 97 A; 166 C; 173 G; 91 T; 0 other;  
 Query Match 70.4%; Score 17.6; DB 21; Length 527;  
 Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 gttgctggtatctgctgtttgaagc 24  
 ||| ||||| ||||| ||||| ||  
 Db 485 GCTGCTGCACGCTGCTTCTTGAAGC 462  
 RESULT 7  
 AAZ53871  
 ID AAZ53871 standard; DNA; 774 BP.  
 XX AC AAZ53871;  
 XX DT 21-MAR-2000 (first entry)  
 XX DE Neisseria meningitidis ORF 568 partial DNA sequence SEQ ID NO:1691.  
 XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;  
 KW antibacterial; gene therapy; ds.  
 XX OS Neisseria meningitidis.  
 XX PN W09957280-A2.  
 XX PD 11-NOV-1999.  
 XX PF 30-APR-1999; 99WO-US09346.  
 XX PR 01-MAY-1998; 98US-0083758.  
 XX PR 31-JUL-1998; 98US-0094869.  
 XX PR 02-SEP-1998; 98US-0098994.  
 XX PR 02-SEP-1998; 98US-0099062.  
 XX PR 09-OCT-1998; 98US-0103749.  
 XX PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 XX  
 DR WPI; 2000-062150/05.  
 DR P-PSDB; AAY75109.  
 XX  
 PT Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics  
 XX  
 PS Claim 7; Page 872; 1453pp; English.  
 XX  
 CC AA253015 to AA254536, AA254577 to AA254615, and AA254616 to AA254617 represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides and polypeptides. AA254537 to AA254576 and AA254616 to AA254617 represent PCR primers used in the exemplification of the present invention. The polypeptides, the polynucleotides, antibodies and compositions of the invention can be used as vaccines, as diagnostic reagents, and as immunogenic compositions. The polypeptides can be used in the manufacture of medicaments for treating or preventing infection due to *Neisseria* bacteria (e.g. meningitis and septicemia), to detect the presence of *Neisseria* bacteria, or to raise antibodies. They may also be used to screen for agonists or antagonists, which may themselves have use as antibacterial agents. The polynucleotides of the invention may also be used in gene therapy protocols.  
 XX  
 SQ Sequence 774 BP; 137 A; 206 C; 210 G; 221 T; 0 other;

Query Match 70.4%; Score 17.6; DB 21; Length 774;  
 Best Local Similarity 83.3%; Pred. No. 1.4e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgctgattctgctgtttgaagcg 25  
 |||| | |||| | |||| |  
 Db 57 ttgcagaattctgcggttgaagcg 80

RESULT 8  
 AAF72295/C  
 ID AAF72295 standard; DNA; 1359 BP.  
 XX  
 AC AAF72295;  
 XX  
 DT 30-APR-2001 (first entry)  
 XX  
 DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:1085.  
 XX  
 KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;  
 KW fine chemical production; microorganism; organic acid; nucleoside;  
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;  
 KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;  
 KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.  
 OS  
 OS Corynebacterium glutamicum.  
 PN  
 PN W0200100843-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 23-JUN-2000; 2000WO-IB00923.  
 XX  
 PR 25-JUN-1999; 99US-0141031.  
 PR 01-JUL-1999; 99DE-1030470.  
 PR 02-JUL-1999; 99US-0142101.  
 PR 08-JUL-1999; 99DE-1031415.  
 PR 08-JUL-1999; 99DE-1031418.

PR 08-JUL-1999; 99DE-1031419.  
 PR 08-JUL-1999; 99DE-1031420.  
 PR 08-JUL-1999; 99DE-1031424.  
 PR 08-JUL-1999; 99DE-1031428.  
 PR 08-JUL-1999; 99DE-1031434.  
 PR 08-JUL-1999; 99DE-1031435.  
 PR 08-JUL-1999; 99DE-1031443.  
 PR 08-JUL-1999; 99DE-1031453.  
 PR 08-JUL-1999; 99DE-1031457.  
 PR 08-JUL-1999; 99DE-1031465.  
 PR 08-JUL-1999; 99DE-1031478.  
 PR 08-JUL-1999; 99DE-1031510.  
 PR 08-JUL-1999; 99DE-1031541.  
 PR 08-JUL-1999; 99DE-1031573.  
 PR 08-JUL-1999; 99DE-1031592.  
 PR 08-JUL-1999; 99DE-1031632.  
 PR 08-JUL-1999; 99DE-1031634.  
 PR 08-JUL-1999; 99DE-1031636.  
 PR 09-JUL-1999; 99DE-1032125.  
 PR 09-JUL-1999; 99DE-1032126.  
 PR 09-JUL-1999; 99DE-1032130.  
 PR 09-JUL-1999; 99DE-1032186.  
 PR 09-JUL-1999; 99DE-1032206.  
 PR 09-JUL-1999; 99DE-1032227.  
 PR 09-JUL-1999; 99DE-1032228.  
 PR 09-JUL-1999; 99DE-1032229.  
 PR 09-JUL-1999; 99DE-1032230.  
 PR 14-JUL-1999; 99DE-1032322.  
 PR 14-JUL-1999; 99DE-1032926.  
 PR 14-JUL-1999; 99DE-1032928.  
 PR 14-JUL-1999; 99DE-1033004.  
 PR 14-JUL-1999; 99DE-1033005.  
 PR 14-JUL-1999; 99DE-1033006.  
 PR 12-AUG-1999; 99US-0148613.  
 PR 27-AUG-1999; 99DE-1040764.  
 PR 27-AUG-1999; 99DE-1040765.  
 PR 27-AUG-1999; 99DE-1040766.  
 PR 27-AUG-1999; 99DE-1040832.  
 PR 31-AUG-1999; 99DE-1041378.  
 PR 31-AUG-1999; 99DE-1041379.  
 PR 31-AUG-1999; 99DE-1041380.  
 PR 31-AUG-1999; 99DE-1041394.  
 PR 31-AUG-1999; 99DE-1041396.  
 PR 03-SEP-1999; 99DE-1042076.  
 PR 03-SEP-1999; 99DE-1042077.  
 PR 03-SEP-1999; 99DE-1042079.  
 PR 03-SEP-1999; 99DE-1042086.  
 PR 03-SEP-1999; 99DE-1042087.  
 PR 03-SEP-1999; 99DE-1042088.  
 PR 03-SEP-1999; 99DE-1042095.  
 PR 03-SEP-1999; 99DE-1042124.  
 PR 03-SEP-1999; 99DE-1042129.  
 PR 09-MAR-2000; 2000US-0187970.  
 XX  
 PA (BADI ) BASF AG.  
 XX  
 PI Pompeius M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
 XX  
 DR WPI; 2001-137957/14.  
 DR P-PSDB; AAB80176.  
 XX  
 XX Nucleic acids from *Corynebacterium glutamicum* encoding metabolic  
 PT pathway proteins, useful for producing fine chemicals in  
 PT microorganisms, including organic acids, nonproteinogenic amino acids,  
 PT and purine and pyrimidine bases -  
 XX  
 XX Claim 3; Page 1637-1639; 1737pp; English.  
 XX  
 CC AAF71753 to AAF72330 encode the *Corynebacterium glutamicum* metabolic  
 CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum  
 CC MP nucleic acids are useful for the production of fine chemicals  
 CC in microorganisms, including organic acids, nonproteinogenic amino  
 CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,





Qy 1 gttgctgagctgctgtttgaagc 24  
 |||| || || ||||| ||||  
 Db 3983 gttgtgaattgctgtttcaagc 4006

## RESULT 11

AAF59633  
 ID AAF59633 standard; cDNA; 4610 BP.  
 XX  
 AC AAF59633;

XX 24-APR-2001 (first entry)

XX Human cell cycle and proliferation protein CCYPR-44 cDNA, SEQ ID NO:98.

XX Cell cycle and proliferation protein; CCYPR; human; agonist;  
 KW antagonist; gene therapy; detection; gene therapy;  
 KW transgenic animal disease model; immune disorder;  
 KW developmental disorder; cell signalling disorder;  
 KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;  
 KW arteriosclerosis; asthma; allergy; diabetes mellitus;  
 KW menstrual cycle disorder; bacterial infection; ss.

XX Homo sapiens.

XX WO200107471-A2.

XX 01-FEB-2001.

XX 21-JUL-2000; 2000WO-US19948.

XX 21-JUL-1999; 99US-0145075.

XX 08-SEP-1999; 99US-0153129.

XX 10-NOV-1999; 99US-0164647.

XX (INCY-) INCYTE GENOMICS INC.

XX Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;

XX Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;

XX WPI; 2001-112727/12.

XX P-PSDB; AAB60496.

XX Human cell cycle and proliferation proteins and polynucleotides are  
 PT used to treat, diagnose and prevent immune, developmental and cell  
 PT signalling disorders and cell proliferative disorders including cancer -

XX Claim 5; Page 197-198; 205pp; English.

XX Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human  
 CC cell cycle and proliferation proteins (CCYPR), AAB60453-AAB60506.  
 CC CCYPR and agonists of CCYPR are used to treat diseases or conditions  
 CC associated with decreased expression of functional CCYPR, while CCYPR  
 CC antagonists are used to treat diseases or conditions associated with  
 CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies  
 CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or  
 CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect  
 CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)  
 CC that specifically bind to CCYPR, and in drug screening methods to  
 CC identify compounds that modulate the activity of CCYPR. CCYPR  
 CC nucleotides can be used to generate transgenic animal models of human  
 CC disease, and can be used in gene therapy in target cells with genetic  
 CC abnormalities with respect to the expression of CCYPR for the  
 CC treatment or prevention of a disorder associated with CCYPR.  
 CC Diseases which can be diagnosed, treated and prevented using CCYPR  
 CC proteins, nucleic acids, agonists or antagonists include immune,  
 CC developmental and cell signalling disorders, and cell proliferative  
 CC disorders including cancer. Specific examples of these disorders  
 CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,  
 CC diabetes mellitus, disorders of the menstrual cycle and infections  
 CC caused by bacteria.

XX Sequence 4610 BP; 1106 A; 1083 C; 1118 G; 1303 T; 0 other;

Query Match 70.4%; Score 17.6; DB 22; Length 4610;  
 Best Local Similarity 83.3%; Pred. No. 1.7e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gttgctgagctgctgtttgaagc 24  
 |||| || || ||||| ||||  
 Db 4015 gttgtgaattgctgtttcaagc 4038

## RESULT 12

AAAB1478/C  
 ID AAAB1478 standard; DNA; 52253 BP.

XX AAAB1478;

XX 04-DEC-2000 (first entry)

XX N. meningitidis partial DNA sequence gnm\_26 SEQ ID NO:26.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; MenB; ds.

XX Neisseria meningitidis.

XX WO200022430-A2.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-US23573.

XX 09-OCT-1998; 98US-0103794.

XX 30-APR-1999; 99US-0132068.

XX (CHIR ) CHIRON CORP.

XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;

XX Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;

XX Rappuoli R, Pizza M;

XX WPI; 2000-318079/27.

XX Isolated nucleotide sequences of Neisseria meningitidis which can be  
 PT used in the diagnosis and treatment of N. meningitidis infection and  
 PT other Neisserial infections, for example, N.gonorrhoea -

XX Claim 7; Page 532-547; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic  
 CC proteins from Neisseria genomic sequences. AAAB1453 to AAAB2414  
 CC represent specifically claimed Neisseria meningitidis genomic DNA  
 CC sequences: AAAB1260 to AAAB1303 and AAB25620 to AAB25663 represent  
 CC Neisseria DNA sequences and their corresponding proteins; AAAB1254 to  
 CC AAAB1259 and AAAB1304 to AAAB1321 represent PCR primers used in the  
 CC isolation of Neisseria meningitidis DNA sequences; and AAAB1322 to  
 CC AAAB1452 represent Neisseria meningitidis MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to  
 CC Neisserial bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against Meningococcus B; against all serotypes;  
 CC and/or against all pathogenic Neisseriae. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and

CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.  
 XX  
 SQ Sequence 52253 BP; 12877 A; 15510 C; 13103 G; 10761 T; 2 other;  
 Query Match 70.4%; Score 17.6; DB 21; Length 52253;  
 Best Local Similarity 83.3%; Pred. No. 2.3e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 ttactgagatctgtgttgagcg 25  
 ||||| ||||| ||||| |||||  
 Db 43436 TTGAGAATCTCGCGGTGAAGCG 43413  
 RESULT 13  
 AAF21544  
 ID AAF21544 standard; DNA; 349980 BP.  
 XX  
 AC AAF21544;  
 XX  
 DT 13-MAR-2001 (first entry)  
 XX  
 DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:1.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;  
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
 KW ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO200066791-A1.  
 XX  
 PD 09-NOV-2000.  
 XX  
 PF 08-MAR-2000; 2000WO-US05928.  
 XX  
 PR 30-APR-1999; 99US-0132068.  
 PR 08-OCT-1999; 99WO-US23573.  
 PR 28-FEB-2000; 2000GB-0004695.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (GENO - ) INST GENOMIC RES.  
 XX  
 PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Massignani V;  
 PI Galsotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;  
 PI Frazer CM, Grandi G;  
 XX  
 WPI; 2000-647603/62.  
 XX  
 PT Neisseria meningitidis B full length genome sequence and open reading  
 PT frames are used to detect, treat and prevent Neisserial infections -  
 XX  
 PS Claim 7; Appendix A; 692pp; English.  
 XX  
 CC The present invention describes the full length genome of  
 CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607  
 CC to AAF21613 represent fragments of the NMB genomic sequence, as the  
 CC sequence was too long to go in a record on its own it was split into 8  
 CC sequences which overlap each other at the beginning and end of each  
 CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at  
 CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at  
 CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the  
 CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to  
 CC AAF21606 represent PCR primers which are used in the exemplification of  
 CC the present invention. The NMB genome and fragments from it have  
 CC antibacterial activity, and can be used in vaccines and gene therapy.  
 CC Neisseria nucleic acids, proteins and/or antibodies which binds to the  
 CC proteins can be used in compositions for treating or preventing infection  
 CC due to Neisserial bacteria or as a diagnostic reagent for detecting the  
 CC presence of Neisserial bacteria or of antibodies raised to Neisserial  
 CC bacteria. Computers, computer memory, computer storage medium or computer  
 CC databases can be used in a search to identify open reading frames (ORFs)

CC or coding sequences within the NMB genome. The DNA sequences provide  
 CC further opportunities to find antigenic or immunogenic proteins which are  
 CC more effective in vaccines than the outer membrane proteins currently  
 CC used.  
 XX  
 SQ Sequence 349980 BP; 83241 A; 85091 C; 95206 G; 86442 T; 0 other;  
 Query Match 70.4%; Score 17.6; DB 21; Length 349980;  
 Best Local Similarity 83.3%; Pred. No. 2.7e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 ttgctgagatctgtgttgagcg 25  
 ||||| ||||| ||||| |||||  
 Db 184668 ttgcagaatctgcggttgagcg 184691  
 RESULT 14  
 AAA81490  
 ID AAA81490 standard; DNA; 1437668 BP.  
 XX  
 AC AAA81490;  
 XX  
 DT 04-DEC-2000 (first entry)  
 XX  
 DE N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; MenB; ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO200022430-A2.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 08-OCT-1999; 99WO-US23573.  
 XX  
 PR 09-OCT-1998; 98US-0103794.  
 PR 30-APR-1999; 99US-0132068.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 XX  
 PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 PI Massignani V, Galsotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
 PI Rappuoli R, Pizza M;  
 XX  
 WPI; 2000-318079/27.  
 XX  
 PT Isolated nucleotide sequences of Neisseria meningitidis which can be  
 PT used in the diagnosis and treatment of N. meningitidis infection and  
 PT other Neisserial infections, for example, N.gonorrhoea -  
 XX  
 PS Claim 7; Page 866-1272; 1760pp; English.  
 XX  
 CC The present invention describes methods of obtaining immunogenic  
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
 CC represent specifically claimed Neisseria meningitidis genomic DNA  
 CC sequences; AAF21260 to AAA81303 and AAB25620 to AAB25663 represent  
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to  
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to  
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to  
 CC Neisserial bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against Meningococcus B; against all serotypes;  
 CC and/or against all pathogenic Neisseriae. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,

CC particularly organism-specific probes. Attempts to make efficacious  
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.

XX SQ Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;

Query Match 70.4%; Score 17.6; DB 21; Length 1437668;  
 Best Local Similarity 83.3%; Pred. No. 2.4e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgctgactctgctgttgagcg 25

Db 184667 ttgcagaatctgcggttgagcg 184690

RESULT 15

AAF21149/c  
 ID AAF21149 standard; DNA; 1994 BP.

XX AC AAF21149;

XX DT 14-MAR-2001 (first entry)

XX DE Human low adenosine antisense oligonucleotide related sequence #2716.

XX KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;  
 KW human; airway disorder; bronchoconstriction; lung inflammation;  
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;  
 KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;  
 KW respiratory obstruction; pulmonary obstruction; impeded respiration;  
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;  
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;  
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;  
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;  
 KW cancer; ss.

XX OS Homo sapiens.

XX PN W0200062736-A2.

XX PD 26-OCT-2000.

XX PF 24-MAR-2000; 2000WO-US08020.

XX PR 06-APR-1999; 99US-0127958.

XX PA (UYEC-) UNIV EAST CAROLINA.

XX PI (NYCE/) NYCE J W.

XX PI Nyce JW;

XX DR WPI; 2000-679539/66.

XX PT Low adenosine (A) content antisense oligonucleotides which do not  
 PT trigger adenosine receptors during metabolism, useful e.g. for treating  
 PT cancers and respiratory obstructions -

XX PS Disclosure; Page 1045-1046; 1592pp; English.

XX CC The present invention describes low adenosine (A) content antisense  
 CC oligonucleotides and compositions (I) comprising them. In the antisense  
 CC oligonucleotides the A is replaced by a 'Universal' or alternative base.  
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,  
 CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.  
 CC The antisense oligonucleotides and (I) can be used to down-regulate the  
 CC expression and/or activity of target polypeptides associated with  
 CC lung/respiratory disorders and malignancies, such as stimulating and

CC activating peptide factors and transmitters, transcription factors,  
 CC immunoglobulins and antibodies, antibody receptors, cytokines and  
 CC chemokines, endogenously produced specific and non-specific enzymes,  
 CC binding proteins, adhesion molecules and their receptors, cytokine and  
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central  
 CC nervous system (CNS) and peripheral nervous and non-nervous system  
 CC receptors, CNS and peripheral nervous and non-nervous system peptide  
 CC transmitters, defensins, growth factors, vasoactive peptides and  
 CC receptors, binding proteins and malignancy associated proteins. The  
 CC antisense oligonucleotides may be used in this way to treat disorders  
 CC including respiratory obstruction (especially pulmonary obstruction  
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)  
 CC and/or surfactant hypoproduction which are associated with a disease or  
 CC condition selected from pulmonary vasoconstriction, inflammation,  
 CC allergies, asthma, impeded respiration, respiratory distress syndrome  
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary  
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),  
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,  
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide  
 CC fragments and antisense oligonucleotides used in the exemplification of  
 CC the present invention.

XX SQ Sequence 1994 BP; 429 A; 585 C; 633 G; 347 T; 0 other;

Query Match 69.6%; Score 17.4; DB 21; Length 1994;

Best Local Similarity 94.7%; Pred. No. 1.9e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 tgctgactctgctgttga 21

||||| |||||||

Db 954 TGCTGGAGCTGCTGTGA 936

RESULT 16

AAC81426/c

ID AAC81426 standard; cDNA; 1994 BP.

XX AC AAC81426;

XX DT 23-FEB-2001 (first entry)

XX DE Human I-kappa-B kinase gamma-subunit (IKK-gamma) cDNA.

XX KW Human; I-kappa-B kinase; IKK; antisense therapy; gene therapy;  
 KW cytokine expression inhibition; NF-kappa-B activation inhibition;  
 KW nuclear factor-kappa-B; rheumatoid arthritis; immune disorder;  
 KW cancer; IKK-gamma; gamma-subunit; ss.

XX OS Homo sapiens.

XX PN JP2000253884-A.

XX PD 19-SEP-2000.

XX PF 10-MAR-1999; 99JP-0063291.

XX PR 10-MAR-1999; 99JP-0063291.

XX PA (TOAG ) TOA GOSEI CHEM IND LTD.

XX PX WPI; 2000-658813/64.

XX PT Antisense nucleic acid compound complementary to the subunit of  
 PT IkappaB, used to treat rheumatic arthritis, immune diseases and cancer

XX PS Claim 3; Page 14-15; 20pp; Japanese.

XX CC The invention relates to an antisense oligonucleotide targetted to  
 CC a gene encoding a subunit of I-kappa-B kinase (IKK) which inhibits its  
 CC expression, and thereby inhibits expression of a cytokine such as  
 CC IL-6 (interleukin-6). I-kappa-B kinase activates NF-kappa-B (nuclear

CC factor-kappa-B) which acts a transcriptional regulator of cytokine  
 CC genes. The antisense oligonucleotide can be used in gene therapy to  
 CC treat rheumatoid arthritis, immune disorders and cancers. Sequences  
 CC AAC81422-C81426 are cDNAs derived from genes whose expression may be  
 CC inhibited using an antisense oligonucleotide of the invention.  
 CC The present sequence represents a human IKK-gamma subunit cDNA.

SQ Sequence 1994 BP; 429 A; 585 C; 633 G; 347 T; 0 other;

Query Match 69.6%; Score 17.4; DB 21; Length 1994;  
 Best Local Similarity 94.7%; Pred. No. 1.9e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 tgctggatctgctgtttga 21  
 ||||| ||||| ||||| |||||  
 Db 954 TGCTGGAGCTGCTGTTTGA 936

RESULT 17  
 ID AAA35027/c  
 AA AAA35027 standard; DNA; 1994 BP.

AC AAA35027;

DT 28-JUL-2000 (first entry)

DE Human adenosine receptor related polynucleotide SEQ ID NO:2716.

KW Human; adenosine receptor; low adenosine antisenase oligonucleotide;  
 KW phosphorothioate; impaired respiration; inflammation; allergy;  
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;  
 KW antiallergic; antiasthmatic; cytotatic; analgesic; impaired airway;  
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;  
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;  
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;  
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

OS Homo sapiens.

XX WO200009525-A2.

PD 24-FEB-2000.

PF 03-AUG-1999; 99WO-US17712.

PR 03-AUG-1998; 98US-0095212.

PA (UYEC-) UNIV EAST CAROLINA.

PI Nyce JW;

DR WPI; 2000-205971/18.

XX New antisense oligonucleotides useful for treating e.g. pulmonary  
 XX vasoconstriction, inflammation, allergies, asthma, hypertension,  
 XX bronchitis, emphysema, respiratory distress syndrome, ischemia or  
 XX cancers -

PS Disclosure; Page 968-969; 1343pp; English.

CC The present invention describes a new composition comprising an  
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which  
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or  
 CC inflammation. The ON can have antiinflammatory, antiallergic,  
 CC antiasthmatic, cytotatic and analgesic activities. The compositions are  
 CC useful for the treatment of diseases associated with inflammation,  
 CC impaired airways, including lung disease and diseases whose secondary  
 CC effects afflict the lungs of a subject. They can be used for treating  
 CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,  
 CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic  
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive  
 CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,

CC carcinomas, and cancers which may metastasise to the lungs, including  
 CC breast and prostate cancer. The reduction of the adenosine content of  
 CC the ONs reduces side effects. The A-containing ONs break down with the  
 CC release of deoxyadenosine which activates adenosine receptors causing  
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the  
 CC nucleotide sequences given in the sequence listing from the present  
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last  
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences  
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680  
 CC (AAA32323 to AAA33992) are specifically claimed ONs from the present  
 CC invention. N.B. Sequences given in the disclosure of the present  
 CC invention do not match up with their corresponding SEQ ID NO: sequences  
 CC given in the sequence listing.

SQ Sequence 1994 BP; 429 A; 585 C; 633 G; 347 T; 0 other;

Query Match 69.6%; Score 17.4; DB 21; Length 1994;  
 Best Local Similarity 94.7%; Pred. No. 1.9e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 tgctggatctgctgtttga 21  
 ||||| ||||| ||||| |||||  
 Db 954 TGCTGGAGCTGCTGTTTGA 936

RESULT 18

AAZ07513/c

ID AAZ07513 standard; DNA; 2009 BP.

XX AAZ07513;

DT 26-NOV-1999 (first entry)

DE Human RIP-associated protein (RAP-2) encoding DNA.

KW Receptor interacting protein; RIP-associated protein-2; RAP-2; RIP;  
 KW inflammation; cell death; cell survival; septic shock; hepatitis;  
 KW graft versus host rejection; diabetes; multiple sclerosis; tumor;  
 KW HIV infection; p55-receptor; FAS-receptor; human; ss.

OS Homo sapiens.

PN WO9947672-A1.

PD 23-SEP-1999.

PF 18-MAR-1999; 99WO-IL00158.

PR 19-MAR-1998; 98IL-0123758.

PR 01-SEP-1998; 98IL-0126024.

PA (YEDA ) YEDA RES & DEV CO LTD.

PA (YESH ) UNIV YESHIVA EINSTEIN COLLEGE.

PI Wallach D, Kovalenko A, Horwitz MS, Li Y;

DR WPI; 1999-562113/47.

DR P-PSDB; AAY27430.

PT New receptor interacting protein-associated protein-2, used to develop  
 PT products for treating, e.g. septic shock, tumors or HIV infection -  
 PS Claim 4; Fig 1A-B; 132pp; English.

CC This DNA encodes a receptor interacting protein (RIP)-associated protein  
 CC -2 (RAP-2). The RAP-2 proteins, isoforms, analogs, fragments or  
 CC derivatives or DNA can be used for the modulation or mediation of the  
 CC RIP modulated/mediated intracellular effects on the inflammation, cell  
 CC death or cell survival pathways in which RIP is involved directly, or  
 CC indirectly via other modulators/mediators of these pathways. They can be  
 CC used for treating e.g. septic shock, graft versus host rejection, acute  
 CC hepatitis, diabetes or multiple sclerosis. They can also be used for

CC treating tumor cells or HIV-infected cells or other diseased cells. The  
 CC RAP-2 binding proteins can be used for modulating/mediating the function  
 CC of RAP-2. The products can also be used for diagnostic purposes, e.g. for  
 CC identifying disorders related to abnormal functioning of cellular effects  
 CC mediated by the p55-R, FAS-R or other related receptors.

XX Sequence 2009 BP; 418 A; 587 C; 643 G; 356 T; 5 other;  
 SQ

Query Match 69.6%; Score 17.4; DB 20; Length 2009;  
 Best Local Similarity 94.7%; Pred. No. 1.9e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 tgctggatcgtgttga 21  
 ||||| |||||  
 Db 959 TGCTGGAGCTGCTGTTGA 941

RESULT 19  
 AAZ07514/C  
 ID AAZ07514 standard; DNA; 2034 BP.  
 XX AC AAZ07514;

XX 26-NOV-1999 (first entry)  
 DT Human RIP-associated protein (RAP-2) clone #41072 nucleotide sequence.

DE Receptor interacting protein; RIP-associated protein-2; RAP-2; RIP;  
 KW inflammation; cell death; cell survival; septic shock; hepatitis;  
 KW graft versus host rejection; diabetes; multiple sclerosis; tumor;  
 KW HIV infection; p55-receptor; FAS-receptor; human; ss.

XX Homo sapiens.  
 OS  
 XX WO9947672-A1.

PN 23-SEP-1999.

XX 18-MAR-1999; 99WO-IL00158.

XX 19-MAR-1998; 98IL-0123758.

PR 01-SEP-1998; 98IL-0126024.

XX (YEDA ) YEDA RES & DEV CO LTD.

PA (YESH ) UNIV YESHIVA EINSTEIN COLLEGE.

XX Wallach D, Kovalenko A, Horwitz MS, Li Y;

XX WPI; 1999-562113/47.

XX New receptor interacting protein-associated protein-2, used to develop  
 PT products for treating, e.g. septic shock, tumors or HIV infection -

XX Claim 5; Fig 2A-B; 132pp; English.

XX The invention relates to receptor interacting protein (RIP)-associated  
 CC protein-2 (RAP-2). The RAP-2 proteins, isoforms, analogs, fragments or  
 CC derivatives or DNA can be used for the modulation or mediation of the  
 CC RIP modulated/mediated intracellular effects on the inflammation, cell  
 CC death or cell survival pathways in which RIP is involved directly, or  
 CC indirectly via other modulators/mediators of these pathways. They can be  
 CC used for treating e.g. septic shock, graft versus host rejection, acute  
 CC hepatitis, diabetes or multiple sclerosis. They can also be used for  
 CC treating tumor cells or HIV-infected cells or other diseased cells. The  
 CC RAP-2 binding proteins can be used for modulating/mediating the function  
 CC of RAP-2. The products can also be used for diagnostic purposes, e.g.  
 CC for identifying disorders related to abnormal functioning of cellular  
 CC effects mediated by the p55-R, FAS-R or other related receptors. The  
 CC present sequence represents the nucleotide sequence of RAP-2 clone  
 CC #41072. This sequence was identical to a 1.5 kb clone (AAZ07513) in its  
 CC coding region, but showed differences in the 5' non-coding region. Both  
 CC these sequences are assumed to be alternatively spliced forms of the same

CC gene.

XX Sequence 2034 BP; 427 A; 588 C; 633 G; 382 T; 4 other;  
 SQ

Query Match 69.6%; Score 17.4; DB 20; Length 2034;  
 Best Local Similarity 94.7%; Pred. No. 1.9e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 tgctggatcgtgttga 21  
 ||||| |||||  
 Db 983 TGCTGGAGCTGCTGTTGA 965

RESULT 20  
 AAF21150/C  
 ID AAF21150 standard; DNA; 8631 BP.

XX AC AAF21150;

XX 14-MAR-2001 (first entry)

DT Human low adenosine antisense oligonucleotide related sequence #2717.

DE Low adenosine antisense oligonucleotide; phosphorothioate; allergy;  
 KW human; airway disorder; bronchoconstriction; lung inflammation;  
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;  
 KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;  
 KW respiratory obstruction; pulmonary obstruction; impeded respiration;  
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;  
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;  
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;  
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;  
 KW cancer; ss.

XX Homo sapiens.

OS WO2000062736-A2.

PN 26-OCT-2000.

XX 24-MAR-2000; 2000WO-US08020.

XX 06-APR-1999; 99US-0127958.

XX (UYEC-) UNIV EAST CAROLINA.

PA (NYCE/) NYCE J W.

XX Nyce JW;

XX WPI; 2000-679539/66.

XX Low adenosine (A) content antisense oligonucleotides which do not  
 PT trigger adenosine receptors during metabolism, useful e.g. for treating  
 PT cancers and respiratory obstructions -

XX Disclosure; Page 1046-1048; 1592pp; English.

XX The present invention describes low adenosine (A) content antisense  
 CC oligonucleotides and compositions (I) comprising them. In the antisense  
 CC oligonucleotides the A is replaced by a 'Universal' or alternative base.  
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,  
 CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.  
 CC The antisense oligonucleotides and (I) can be used to down-regulate the  
 CC expression and or activity of target polypeptides associated with  
 CC lung/respiratory disorders and malignancies, such as stimulating and  
 CC activating peptide factors and transmitters, transcription factors,  
 CC immunoglobulins and antibodies, antibody receptors, cytokines and  
 CC chemokines, endogenously produced specific and non-specific enzymes,  
 CC binding proteins, adhesion molecules and their receptors, cytokine and  
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central  
 CC nervous system (CNS) and peripheral nervous and non-nervous system  
 CC receptors, CNS and peripheral nervous and non-nervous system peptide

transmitters, defensins, growth factors, vasoactive peptides and receptors, binding proteins and malignancy associated proteins. The antisense oligonucleotides may be used in this way to treat disorders including respiratory obstruction (especially pulmonary obstruction and/or bronchoconstriction) and/or lung inflammation, allergies (and/or surfactant hypoproduction which are associated with a disease or condition selected from pulmonary vasoconstriction, inflammation, allergies, asthma, impaired respiration, respiratory distress syndrome (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), pulmonary transplantation rejection, pulmonary infections, bronchitis, and/or cancer. AAF19434 to AAF21543 represent human polynucleotide fragments and antisense oligonucleotides used in the exemplification of the present invention.

Sequence 8631 BP; 2280 A; 2050 C; 2270 G; 2031 T; 0 other;

Query Match 69.6%; Score 17.4; DB 21; Length 8631;  
Best Local Similarity 94.7%; Pred. No. 2.3e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 tgcgtgacgtcgtgttga 21  
||||| |||||||

Db 7591 TGCTGGAGCTGCTGTGA 7573

RESULT 21  
AAA35028/C  
ID AAA35028 standard; DNA; 8631 BP.  
XX  
AC AAA35028;  
XX  
DT 28-JUL-2000 (first entry)  
XX  
DE Human adenosine receptor related polynucleotide SEQ ID NO:2717.  
XX  
KW Human; adenosine receptor; low adenosine antisense oligonucleotide; phosphorothioate; impaired respiration; inflammation; allergy;  
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;  
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;  
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;  
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;  
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;  
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200009525-A2.  
XX  
PD 24-FEB-2000.  
XX  
PF 03-AUG-1999; 99WO-US17712.  
XX  
PR 03-AUG-1998; 98US-0095212.  
XX  
PA (UYEC-) UNIV EAST CAROLINA.  
XX  
PI Nyce JW;  
XX  
DR WPI; 2000-205971/18.  
XX  
PT New antisense oligonucleotides useful for treating e.g. pulmonary  
PT vasoconstriction, inflammation, allergies, asthma, hypertension,  
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or  
PT cancers -  
XX  
PS Disclosure; Page 969-971; 1343pp; English.  
XX  
CC The present invention describes a new composition comprising an  
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which  
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or  
CC inflammation. The ON can have antiinflammatory, antiallergic,

antiasthmatic, cytostatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation, impaired airways, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma, impaired respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukaemias, lymphomas, carcinomas, and cancers which may metastasize to the lungs, including breast and prostate cancer. The reduction of the adenosine content of the ONs reduces side effects. The A-containing ONs break down with the release of deoxyadenosine which activates adenosine receptors causing bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the nucleotide sequences given in the sequence listing from the present invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185 sequences are also called SEQ ID NO:1 to 185, but the sequences differ from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to AAA33992) are specifically claimed ONs from the present invention. N.B. Sequences given in the disclosure of the present invention do not match up with their corresponding SEQ ID NO: sequences given in the sequence listing.

Sequence 8631 BP; 2280 A; 2050 C; 2270 G; 2031 T; 0 other;

Query Match 69.6%; Score 17.4; DB 21; Length 8631;  
Best Local Similarity 94.7%; Pred. No. 2.3e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 tgcgtgacgtcgtgttga 21  
||||| |||||||

Db 7591 TGCTGGAGCTGCTGTGA 7573

RESULT 22  
AAC03560/C  
ID AAC03560 standard; cDNA; 434 BP.  
XX  
AC AAC03560;  
XX  
DT 06-OCT-2000 (first entry)  
XX  
DE Human secreted protein 5' EST, SEQ ID NO: 3558.  
XX  
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1033401-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 21-FEB-2000; 2000EP-0200610.  
XX  
PR 26-FEB-1999; 99US-0122487.  
XX  
PA (GEST ) GENSET.  
XX  
PI Dumas Milne Edwards J, Duclert A, Giordano J;  
XX  
DR WPI; 2000-500381/45.  
XX  
PT P-PSDB; AAG03554.  
XX  
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
PS Claim 1; SEQ ID 3558; 71pp + CD-ROM; English.  
XX  
CC The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. An ORF has been identified within the  
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs

CC derived from 30 different tissues. EST sequences usually correspond  
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
 CC well suited for isolating cDNA sequences derived from the 5' ends of  
 CC mRNAs and even in those cases where longer cDNA sequences have been  
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
 CC gene therapy and chromosome mapping procedures. They are used to obtain  
 CC upstream regulatory sequences and to design expression and secretion  
 CC vectors.  
 XX  
 SQ Sequence 434 BP; 136 A; 73 C; 101 G; 122 T; 2 other;

Query Match 68.8%; Score 17.2; DB 21; Length 434;  
 Best Local Similarity 86.4%; Pred. No. 1.9e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 tctgtgatctgtgttgagc 24  
 || ||||| ||||| |||||  
 Db 297 TGTGGATCTCTGTTGAACC 276

RESULT 23  
 AAF13715/C  
 ID AAF13715 standard; cDNA; 687 BP.

XX AAF13715;

AC AAF13715;

DT 13-MAR-2001 (first entry)

XX Aspergillus oryzae EST SEQ ID NO:6238.

XX Multiple gene expression; filamentous fungal cell; EST;

KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;

KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;

KW culture condition; environmental stress; spore morphogenesis;

KW metabolic pathway engineering; catabolic pathway engineering; ss.

XX Aspergillus oryzae.

XX WQ200056762-A2.

PN 28-SEP-2000.

XX 22-MAR-2000; 2000WO-US07781.

XX 22-MAR-1999; 99US-0273623.

XX (NOVO ) NOVO NORDISK BIOTECH INC.

PA (NOVO ) NOVO NORDISK AS.

XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

XX WPI; 2000-594572/56.

XX Monitoring differential expression of genes in filamentous fungal cells

XX uses fluorescence-labeled nucleic acids isolated from the cells and a

XX substrate of expressed sequence tags -

XX Claim 88; Page 2566; 3161pp; English.

XX The present invention describes a method for monitoring differential

XX expression of genes in a first filamentous fungal (FF) cell relative to

XX expression of the same genes in one or more second filamentous fungal

XX cells. The method uses fluorescence-labeled nucleic acids isolated from

XX the FF cells and a substrate of expressed sequence tags (EST). The ESTs

XX are used in the methods for monitoring differential expression of genes

XX in a first filamentous fungal (FF) cell relative to expression of the

XX same genes in one or more second filamentous fungal cells. Monitoring

XX the global expression of genes from FF cells allows the production

XX potential of the microorganisms to be improved. New genes may be

CC discovered, possible functions of unknown open reading frames can be  
 CC identified and gene copy number variation and stability can be  
 CC monitored. The expression of genes can be used to study how FF cells  
 CC adapt to changes in culture conditions, environmental stress, spore  
 CC morphogenesis, recombination, metabolic or catabolic pathway  
 CC engineering. Using ESTs provides several advantages over genomic or  
 CC random cDNA clones including elimination of redundancy as one spot on an  
 CC array equals one gene or open reading frame, and organisation of the  
 CC microarrays based on function of the gene products to facilitate  
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from  
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus  
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and  
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
 CC all specifically claimed in the present invention.  
 XX  
 SQ Sequence 687 BP; 185 A; 202 C; 166 G; 129 T; 5 other;

Query Match 68.8%; Score 17.2; DB 21; Length 687;  
 Best Local Similarity 86.4%; Pred. No. 2e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gttgtgatctgtgtttgaa 22  
 || ||||| ||||| |||||  
 Db 565 GTCGCTGGAACTGCTGATTGAA 544

RESULT 24

AAZ95242/C

ID AAZ95242 standard; DNA; 1068 BP.

XX AAZ95242;

AC AAZ95242;

DT 05-JUN-2000 (first entry)

XX Cyclophilin-type peptidyl prolyl cis/trans isomerase nucleotide sequence.

DE Cyclophilin-type peptidyl prolyl cis/trans isomerase; CPC1; cancer; AIDS;

KW leukaemia; reproductive disorder; asthma; diabetes; infertility; anaemia;

KW polycystic ovary syndrome; uterine fibroid; Good pasture's syndrome;

KW gout; Grave's disease; multiple sclerosis; lupus; osteoarthritis;

KW irritable bowel syndrome; ds.

XX Homo sapiens.

OS US6030825-A.

XX 29-FEB-2000.

XX 19-AUG-1998; 98US-0136442.

XX 19-AUG-1998; 98US-0136442.

XX (INCY-) INCYTE PHARM INC.

XX Hillman JL, Corley NC, Patterson C, Guegler KJ;

XX WPI; 2000-205207/18.

XX P-PSDB; AAY78941.

XX Isolated polynucleotides encoding cyclophilin-type peptidyl-prolyl

XX cis/trans isomerase, useful for preventing, diagnosing and treating

XX cancers, autoimmune/inflammatory disorders and reproductive diseases -

XX Claim 1; Fig 1; 28pp; English.

XX This sequence represents a human cyclophilin-type peptidyl-prolyl

XX cis/trans isomerase (CPC1) nucleotide sequence. The invention includes

XX probes for the CPC1 nucleotide sequence and vectors expressing the

XX polynucleotide. CPC1 is a member of the peptidyl/prolyl cis/trans

XX isomerase (PPIase) class of enzymes. Cyclophilin isomerase activity is

XX essential for correct protein folding and protein trafficking. The CPC1

XX nucleotide sequence and the protein it encodes may be used in the

CC diagnosis, prevention and treatment of disorders associated with  
 CC inappropriate CPCI expression and activity. For example, they may be used  
 CC to treat cancers (e.g. leukaemia, lymphoma, melanoma and cancers of the  
 CC breast, liver and prostate), autoimmune/inflammatory disorders  
 CC (e.g. AIDS, asthma and diabetes mellitus) and reproductive disorders  
 CC (e.g. infertility, polycystic ovary syndrome and uterine fibroids). The  
 CC nucleotide sequence may be used to treat and diagnose allergy,  
 CC anaemia, Goodpasture's syndrome, Crohn's disease, gout, Grave's disease,  
 CC multiple sclerosis, lupus, irritable bowel syndrome, ulcerative colitis  
 CC and osteoarthritis. The CPCI polynucleotide or vectors containing it may  
 CC be administered to treat any of the above diseases by rectifying  
 CC mutations or deletions in a patient's genome, affecting CPCI metabolism  
 CC by expressing inactive proteins or to supplement the patients' own  
 CC production of CPCI proteins. Conversely, antisense nucleic acid molecules  
 CC may be administered to down regulate CPCI protein expression by binding  
 CC with the cells own CPCI genes and preventing their expression. Sense and  
 CC antisense CPCI nucleotide sequences may also be used as DNA probes in  
 CC diagnostic assays (e.g. PCR) to detect and quantitate the presence of  
 CC similar nucleic acid sequences in samples, and hence which patients may  
 CC be in need of restorative therapy. They may also be used to study the  
 CC expression and function of CPCI protein domains and their role in  
 CC metabolism.

SQ Sequence 1068 BP; 341 A; 200 C; 206 G; 321 T; 0 other;

Query Match 68.8%; Score 17.2; DB 21; Length 1068;  
 Best Local Similarity 86.4%; Pred. No. 2.1e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 tgcgtgacgtcgtgttggaac 24  
 || ||||| ||||| |||||  
 Db 363 TGTTGGATCTCTGTTGAAC 342

RESULT 25

AAT03854/c  
 ID AAT03854 standard; DNA; 10564 BP.

XX AC AAT03854;

XX DT 26-OCT-1996 (first entry)

XX DE Human fibrinogen gamma chain coding sequence.

XX KW Human fibrinogen gamma chain; transgenic animal milk; treatment;  
 KW sheep beta-lactoglobulin signal peptide fusion protein;  
 KW surgical adhesive; ds.

XX OS Homo sapiens.

Key	Location/Qualifiers
FT 5'UTR	1..1798
FT exon	/*tag= a 1799..1876
FT intron	/*tag= b /*note= "1" 1877..1972
FT exon	/*tag= c /*note= "1" 1973..2017
FT intron	/*tag= d /*note= "2" 2018..2206
FT exon	/*tag= e /*note= "2" 2207..2390
FT intron	/*tag= f /*note= "3" 2391..2509
FT exon	/*tag= g /*note= "3" 2510..2603

FT intron	/*tag= h /*note= "4" 2604..4210
FT exon	/*tag= i /*note= "4" 4211..4341
FT intron	/*tag= j /*note= "5" 4342..4644
FT exon	/*tag= k /*note= "5" 4645..4778
FT intron	/*tag= l /*note= "6" 4779..5757
FT exon	/*tag= m /*note= "6" 5758..5942
FT intron	/*tag= n /*note= "7" 5943..7425
FT exon	/*tag= o /*note= "7" 7426..7703
FT intron	/*tag= p /*note= "8" 7704..9341
FT exon	/*tag= q /*note= "8" 9342..9574
FT 3'UTR	/*tag= r /*note= "9" 9575..10564
FT CDS	/*tag= s 1799-9574
FT	/*tag= t /*note= "exons 1-9"
XX WO9523868-A1.	
XX 08-SEP-1995.	
XX 01-MAR-1995;	95WO-US02648.
XX 03-MAR-1994;	94US-0206176.
XX (PHAR-) PHARM PROTEINS LTD.	
XX (ZYMO) ZYMOGENETICS INC.	
XX Dalrymple MA, Foster DC, Garner I, Prunkard DE;	
XX WPI; 1995-320582/41.	
XX P-PSDB; AAR82245.	
XX Production of fibrinogen in transgenic mammals - by introducing DNA segments into the germ line of a non-human mammal and collecting milk from female progeny.	
XX Disclosure; Page 57-67; 99pp; English.	
XX This sequence encodes the human fibrinogen gamma chain. It may be operably linked to the signal peptide, preferably of the sheep beta-lactoglobulin gene (see AAT03855) and, together with the fibrinogen A-alpha chain sequence (see AAT03853) and the B-beta chain sequence (see AAT03852), is introduced into a fertilised egg or the germ line of a non-human animal, preferably a sheep, pig, goat or cattle at a molar ratio of 0.5:1.0:0.5:1.0:0.5:1.0, respectively, for fusion protein gene expression in transgenic animal milk. The recombinantly produced active fibrinogen is useful in human and veterinary medicine, e.g. in the formulation of surgical adhesives, which also consist of Factor-XIII, and as a coating surface for polymeric articles, such as synthetic vascular grafts.	







PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 68.0%; Score 17; DB 21; Length 492;  
 Best Local Similarity 80.0%; Pred. No. 2.3e+02;  
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gttgctggaatctgctgtttgaagcg 25  
 ||||| ||||| ||||| |||||  
 Db 242 gatgctgggtctgctggtttaag 266

RESULT 31  
 AAV87783  
 ID AAV87783 standard; cDNA; 729 BP.  
 XX  
 AC AAV87783;

DT 12-FEB-1999 (first entry)

DE EST clone ES306.

XX Expressed sequence tag; secreted protein; haematopoiesis regulator;  
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

XX Homo sapiens.

OS

XX

PN WO9845437-A2.

XX

PD 15-OCT-1998.

XX

PF 10-APR-1998; 98WO-US06956.

XX

PR 10-APR-1997; 97US-0837312.

XX

PA (GEMY ) GENETICS INST INC.

XX

XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

PI Racie LA, Spaulding V, Treacy M;

XX

DR WPI; 1999-070078/06.

XX

PT New polynucleotides encoding human secreted proteins - derived from

PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,

PT ovary, pituitary, retina and colon cDNA libraries

XX

PS Claim 1; Page 178; 641pp; English.

XX

CC The present sequence represents an expressed sequence tag (EST), and is

CC a polynucleotide of the invention. The polynucleotides of the invention

CC are all secreted EST sequences isolated from a variety of human tissue

CC sources. The EST sequences and proteins encoded by them are predicted to

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 gttgctggaatctgctgtttgaagcg 25  
 ||||| ||||| ||||| |||||  
 Db 532 gttgtaggtctgctgttggtagcg 556

RESULT 32  
 AAA58914  
 ID AAA58914 standard; DNA; 751 BP.  
 XX  
 AC AAA58914;

DT 20-OCT-2000 (first entry)

DE DNA encoding ketosynthase domain of epothilone polyketide synthase.

XX

DE Epothilone polyketide synthase; PKS; ketosynthase; acyltransferase;

KW acyl carrier protein; polyketide; ketoreductase; enoylreductase;

KW dehydratase; epothilone; antibiotic; antitumor; ss.

XX

OS Sorangium cellulosum.

XX

XX Key Location/Qualifiers

FT CDS

FT 1..751

FT /\*tag= a

FT /transl\_except= (pos: 124..126, aa: Xaa)

FT /note= "Xaa is any amino acid"

XX

PN US090601-A.

XX

PD 18-JUL-2000.

XX

PF 23-JAN-1998; 98US-0010809.

XX

PR 23-JAN-1998; 98US-0010809.

XX

PA (KOSA-) KOSAN BIOSCIENCE.

XX

XX Gustafsson C, Betlach MC;

XX

XX WPI; 2000-498261/44.

DR P-PSDB; AAB07517.

XX

PT Novel epothilone polyketide synthase comprising a ketoreductase or

PT ketosynthase domain, useful for producing epothilone -

XX

PS Claim 6; Column 15-16; 39pp; English.

XX

CC The present sequence encodes a ketosynthase domain of epothilone

CC polyketide synthase (PKS). In type I or modular PKS enzymes, a set of

Query Match 68.0%; Score 17; DB 21; Length 751;  
 Best Local Similarity 80.0%; Pred. No. 2.5e+02;  
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 gttgctggaatctgctgtttgaagcg 25  
 ||||| ||||| ||||| |||||  
 Db 222 gttgctgaagcgctgctgtagcg 246

RESULT 32  
 AAA58914  
 ID AAA58914 standard; DNA; 751 BP.  
 XX  
 AC AAA58914;

DT 20-OCT-2000 (first entry)

DE DNA encoding ketosynthase domain of epothilone polyketide synthase.

XX

DE Epothilone polyketide synthase; PKS; ketosynthase; acyltransferase;

KW acyl carrier protein; polyketide; ketoreductase; enoylreductase;

KW dehydratase; epothilone; antibiotic; antitumor; ss.

XX

OS Sorangium cellulosum.

XX

XX Key Location/Qualifiers

FT CDS

FT 1..751

FT /\*tag= a

FT /transl\_except= (pos: 124..126, aa: Xaa)

FT /note= "Xaa is any amino acid"

XX

PN US090601-A.

XX

PD 18-JUL-2000.

XX

PF 23-JAN-1998; 98US-0010809.

XX

PR 23-JAN-1998; 98US-0010809.

XX

PA (KOSA-) KOSAN BIOSCIENCE.

XX

XX Gustafsson C, Betlach MC;

XX

XX WPI; 2000-498261/44.

DR P-PSDB; AAB07517.

XX

PT Novel epothilone polyketide synthase comprising a ketoreductase or

PT ketosynthase domain, useful for producing epothilone -

XX

PS Claim 6; Column 15-16; 39pp; English.

XX

CC The present sequence encodes a ketosynthase domain of epothilone

CC polyketide synthase (PKS). In type I or modular PKS enzymes, a set of



DE Arabidopsis thaliana DNA fragment SEQ ID NO: 59221.  
XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX EP1033405-A2.  
PN  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.

```

PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0158065.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match      68.0%; Score 17; DB 21; Length 1200;
Best Local Similarity 80.0%; Pred. No. 2.6e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gttgctgagctgctgtttgaacg 25
Db 200 gatgctgggtctgctggtttaag 224

RESULT 36
AAC34154
ID AAC34154 standard; DNA; 1237 BP.
XX AC AAC34154;
XX XX
XX 17-OCT-2000 (first entry)
XX XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 5643.
XX XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX XX
XX Arabidopsis thaliana.

```

```

XX XX
PN PN
XX XX EP1033405-A2.
PD PD 06-SEP-2000.
XX XX
PF PF 25-FEB-2000; 2000EP-0301439.
XX XX
PR PR 25-FEB-1999; 99US-0121825.
PR PR 05-MAR-1999; 99US-0123180.
PR PR 09-MAR-1999; 99US-0123548.
PR PR 23-MAR-1999; 99US-0125788.
PR PR 25-MAR-1999; 99US-0126264.
PR PR 29-MAR-1999; 99US-0126785.
PR PR 01-APR-1999; 99US-0127462.
PR PR 06-APR-1999; 99US-0128234.
PR PR 08-APR-1999; 99US-0128714.
PR PR 16-APR-1999; 99US-0129845.
PR PR 19-APR-1999; 99US-0130077.
PR PR 21-APR-1999; 99US-0130449.
PR PR 23-APR-1999; 99US-0130510.
PR PR 23-APR-1999; 99US-0130891.
PR PR 30-APR-1999; 99US-0131449.
PR PR 30-APR-1999; 99US-0132048.
PR PR 04-MAY-1999; 99US-0132407.
PR PR 05-MAY-1999; 99US-0132484.
PR PR 06-MAY-1999; 99US-0132485.
PR PR 06-MAY-1999; 99US-0132486.
PR PR 07-MAY-1999; 99US-0132863.
PR PR 11-MAY-1999; 99US-0134256.
PR PR 14-MAY-1999; 99US-0134218.
PR PR 14-MAY-1999; 99US-0134219.
PR PR 14-MAY-1999; 99US-0134221.
PR PR 14-MAY-1999; 99US-0134370.
PR PR 18-MAY-1999; 99US-0134768.
PR PR 19-MAY-1999; 99US-0134941.
PR PR 20-MAY-1999; 99US-0135124.
PR PR 21-MAY-1999; 99US-0135353.
PR PR 24-MAY-1999; 99US-0135629.
PR PR 25-MAY-1999; 99US-0136021.
PR PR 27-MAY-1999; 99US-0136392.
PR PR 28-MAY-1999; 99US-0136782.
PR PR 01-JUN-1999; 99US-0137222.
PR PR 03-JUN-1999; 99US-0137528.
PR PR 04-JUN-1999; 99US-0137502.
PR PR 07-JUN-1999; 99US-0137724.
PR PR 08-JUN-1999; 99US-0138094.
PR PR 10-JUN-1999; 99US-0138540.
PR PR 10-JUN-1999; 99US-0138847.
PR PR 14-JUN-1999; 99US-0139119.
PR PR 16-JUN-1999; 99US-0139452.
PR PR 16-JUN-1999; 99US-0139453.
PR PR 17-JUN-1999; 99US-0139492.
PR PR 18-JUN-1999; 99US-0139454.
PR PR 18-JUN-1999; 99US-0139455.
PR PR 18-JUN-1999; 99US-0139456.
PR PR 18-JUN-1999; 99US-0139457.
PR PR 18-JUN-1999; 99US-0139458.
PR PR 18-JUN-1999; 99US-0139459.
PR PR 18-JUN-1999; 99US-0139460.
PR PR 18-JUN-1999; 99US-0139461.
PR PR 18-JUN-1999; 99US-0139462.
PR PR 18-JUN-1999; 99US-0139463.
PR PR 18-JUN-1999; 99US-0139750.
PR PR 18-JUN-1999; 99US-0139763.
PR PR 21-JUN-1999; 99US-0139817.
PR PR 22-JUN-1999; 99US-0139899.
PR PR 23-JUN-1999; 99US-0140353.
PR PR 23-JUN-1999; 99US-0140354.
PR PR 24-JUN-1999; 99US-0140695.
PR PR 28-JUN-1999; 99US-0140823.
PR PR 29-JUN-1999; 99US-0140991.
PR PR 30-JUN-1999; 99US-0141287.

```







PR	16-JUL-1999;	99US-01144088;
PR	16-JUL-1999;	99US-01144088;
PR	16-JUL-1999;	99US-01144325;
PR	19-JUL-1999;	99US-01144331;
PR	19-JUL-1999;	99US-01144332;
PR	19-JUL-1999;	99US-01144333;
PR	19-JUL-1999;	99US-01144334;
PR	19-JUL-1999;	99US-01144335;
PR	20-JUL-1999;	99US-01144352;
PR	20-JUL-1999;	99US-01144632;
PR	21-JUL-1999;	99US-01144884;
PR	21-JUL-1999;	99US-01144814;
PR	21-JUL-1999;	99US-01145086;
PR	21-JUL-1999;	99US-01145088;
PR	22-JUL-1999;	99US-01145088;
PR	22-JUL-1999;	99US-01145087;
PR	22-JUL-1999;	99US-01145089;
PR	22-JUL-1999;	99US-01145192;
PR	23-JUL-1999;	99US-01145144;
PR	23-JUL-1999;	99US-01145218;
PR	23-JUL-1999;	99US-01145224;
PR	26-JUL-1999;	99US-01145276;
PR	27-JUL-1999;	99US-01145913;
PR	27-JUL-1999;	99US-01145918;
PR	27-JUL-1999;	99US-01145919;
PR	28-JUL-1999;	99US-01145951;
PR	02-AUG-1999;	99US-01146386;
PR	02-AUG-1999;	99US-01146388;
PR	02-AUG-1999;	99US-01146389;
PR	03-AUG-1999;	99US-01147038;
PR	04-AUG-1999;	99US-01147204;
PR	05-AUG-1999;	99US-01147302;
PR	05-AUG-1999;	99US-01147192;
PR	06-AUG-1999;	99US-01147303;
PR	06-AUG-1999;	99US-01147416;
PR	09-AUG-1999;	99US-01147493;
PR	10-AUG-1999;	99US-01147935;
PR	10-AUG-1999;	99US-01148171;
PR	11-AUG-1999;	99US-01148319;
PR	12-AUG-1999;	99US-01148341;
PR	13-AUG-1999;	99US-01148565;
PR	13-AUG-1999;	99US-01149368;
PR	16-AUG-1999;	99US-01149175;
PR	17-AUG-1999;	99US-01149426;
PR	18-AUG-1999;	99US-01149722;
PR	20-AUG-1999;	99US-01149723;
PR	20-AUG-1999;	99US-01149929;
PR	23-AUG-1999;	99US-01149902;
PR	23-AUG-1999;	99US-01149930;
PR	25-AUG-1999;	99US-01150566;
PR	26-AUG-1999;	99US-01150884;
PR	27-AUG-1999;	99US-01151065;
PR	27-AUG-1999;	99US-01151066;
PR	27-AUG-1999;	99US-01151080;
PR	30-AUG-1999;	99US-01151303;
PR	31-AUG-1999;	99US-01151438;
PR	01-SEP-1999;	99US-01151930;
PR	07-SEP-1999;	99US-01523363;
PR	10-SEP-1999;	99US-0153070;
PR	13-SEP-1999;	99US-01537558;
PR	15-SEP-1999;	99US-01540118;
PR	16-SEP-1999;	99US-0154039;
PR	20-SEP-1999;	99US-0154779;
PR	22-SEP-1999;	99US-01551339;
PR	23-SEP-1999;	99US-0155486;
PR	24-SEP-1999;	99US-0155659;
PR	28-SEP-1999;	99US-0156458;
PR	29-SEP-1999;	99US-0156596;
PR	04-OCT-1999;	99US-0157117;
PR	05-OCT-1999;	99US-01577553;
PR	06-OCT-1999;	99US-0157865;

PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	23-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match 68.0%; Score 17; DB 21; Length 1573;  
Best Local Similarity 80.0%; Pred. No. 2.7e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels

QY      1 gttgctggatctgctgttgaagcg 25  
         | ||||| ||||| | ||| |  
Db    517 gatgctgggtctgctggtttaagaq 541

RESULT	39	
AAC	37600	
ID	AAC37600 standard; DNA; 1574 BP.	
XX		
XX	AAC37600;	
XX		
DT	17-OCT-2000 (first entry)	
XX		
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 17980.	
XX		
KW	Hybridisation assay; genetic mapping; gene expression control;	
KW	protein identification; signal transduction pathway;	
KW	metabolic pathway; promoter; termination sequence; ss.	
XX		
OS	Arabidopsis thaliana.	
XX		
PN	EP1033405-A2.	
XX		
PD	06-SEP-2000.	
XX		
PF	25-FEB-2000; 2000EP-0301439.	
XX		
PR	25-FEB-1999; 99US-0121825.	
PR	05-MAR-1999; 99US-0123180.	
PR	09-MAR-1999; 99US-0123548.	
PR	23-MAR-1999; 99US-0125788.	
PR	25-MAR-1999; 99US-0126264.	
PR	29-MAR-1999; 99US-0126785.	
PR	01-APR-1999; 99US-0127462.	
PR	06-APR-1999; 99US-0128234.	
PR	18-APR-1999; 99US-0128714.	
PR	16-APR-1999; 99US-0129845.	



PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 68.0%; Score 17; DB 21; Length 1574;  
 Best Local Similarity 80.0%; Pred. No. 2.7e+02;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gttgctggatctgctgtttgaagcg 25  
 | ||||| ||||| || |||||  
 Db 516 gatgctggctgctgtgttaagag 540

RESULT 40

AAC88117  
 ID AAC88117 standard; cDNA; 2750 BP.

AC AAC88117;

DT 09-MAR-2001 (first entry)

DE Human FLEXHT-48 nucleotide sequence SEQ ID NO:103.

Human: FLEXHT: full-length molecules expressed in human tissue;  
 diagnosis; gene expression; genetic linkage; genetic variability;  
 antianemic; anticonvulsant; antiarteriosclerotic; immunomodulatory;  
 cytotatic; antiasthmatic; antiinflammatory; hepatotropic; antidiabetic;  
 anti-gout; antithyroid; neuroprotective; antiarthritic; osteopathic;  
 antipsoriatic; antirheumatic; antiulcer; gene therapy; anaemia; gout;  
 epilepsy; arteriosclerosis; atherosclerosis; developmental disorder;  
 cancer; immunological disorder; asthma; bronchitis; cirrhosis;  
 Crohn's disease; diabetes mellitus; Grave's disease; multiple sclerosis;  
 osteoarthritis; pancreatitis; rheumatoid arthritis; psoriasis;  
 ulcerative colitis; ss.

OS Homo sapiens.

PN WO200070047-A2.

PD 23-NOV-2000.

PF 12-MAY-2000; 2000WO-US13299.

PR 14-MAY-1999; 99US-0311894.

PR 14-MAY-1999; 99US-0311937.

PR 14-MAY-1999; 99US-0311940.

XX (INCY-) INCYTE GENOMICS INC.

XX Yue H, Tang YT, Lal P, Reddy R, Batra S, Baughn MR, Yang J;

PI Azimzai Y, Lu DAM, Au-Young J, Shih LL;

XX

DR WPI; 2001-016234/02.  
 DR P-PSDB; AAB36626.  
 XX Human FLEXHT protein and DNA sequences, useful for treating  
 PT immunological disorders, developmental disorders, and cancers -  
 XX Claim 5; Page 163-164; 168pp; English.

XX AAC88070 to AAC88124 encode the 55 FLEXHT (full-length molecules  
 CC expressed in human tissues) proteins given in AAB36579 to AAB36633. The  
 CC present invention describes an isolated polypeptide (A) comprising an  
 CC amino acid sequence selected from one of 55 amino acid sequences 42-876  
 CC residues in length, corresponding to FLEXHT-1 to FLEXHT-55, a 90 %  
 CC identical sequence, and a biologically active or immunogenic fragment of  
 CC the sequence. The FLEXHT proteins can have antianemic, anticonvulsant,  
 CC antiarteriosclerotic, immunomodulatory, cytostatic, antiasthmatic,  
 CC antiinflammatory, hepatotropic, antidiabetic, anti-gout, antithyroid,  
 CC neuroprotective, antiarthritic, osteopathic, antipsoriatic, antiulcer  
 CC and antirheumatic activities, and can be used in gene therapy. The  
 CC polynucleotide sequences can be used to express the protein sequences.  
 CC Pharmaceutical compositions comprising FLEXHT can be used to treat  
 CC diseases or conditions associated with altered expression of functional  
 CC FLEXHT. The proteins and polynucleotides can be used to diagnose and  
 CC treat disorders including anaemia, epilepsy, arteriosclerosis,  
 CC atherosclerosis, developmental disorders, cancers, and immunological  
 CC disorders such as asthma, bronchitis, cirrhosis, Crohn's disease,  
 CC diabetes mellitus, gout, Grave's disease, multiple sclerosis,  
 CC osteoarthritis, pancreatitis, psoriasis, rheumatoid arthritis, and  
 CC ulcerative colitis.

XX Sequence 2750 BP; 776 A; 619 C; 638 G; 717 T; 0 other;

Query Match 68.0%; Score 17; DB 22; Length 2750;

Best Local Similarity 80.0%; Pred. No. 2.9e+02;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gttgctggatctgctgtttgaagcg 25  
 |||| | ||||| || |||||  
 Db 789 gttgtaggtctgctgtgtgtagcg 813

RESULT 41

AAT35869

ID AAT35869 standard; cDNA; 3435 BP.

XX AAT35869;

XX 21-OCT-1996 (first entry)

XX Human DNA polymerase delta cDNA.

XX DNA polymerase delta; pol delta; colorectal tumour; cancer;  
 KW gene therapy; diagnosis; replication error; RER; DNA repair; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 43..3366

FT /\*tag= a

XX WO9621026-A1.

XX 11-JUL-1996.

XX 02-JAN-1996; 96WO-US000005.

XX 30-DEC-1994; 94US-0366577.

XX (UYJO ) UNIV JOHNS HOPKINS.

XX Kinzler KW, Vogelstein B;

XX



CC therapy.  
 XX Sequence 309 BP; 87 A; 86 C; 64 G; 72 T; 0 other;  
 SQ

Query Match 67.2%; Score 16.8; DB 20; Length 309;  
 Best Local Similarity 90.0%; Pred. No. 2.7e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 tgctggatctgctgtttgaa 22  
 ||||| ||||| |||||  
 Db 44 tgctgtatctgctgtttgaa 63

RESULT 44  
 AAV58754  
 ID AAV58754 standard; cDNA; 1328 BP.  
 XX  
 AC AAV58754;  
 XX  
 DT 18-JAN-1999 (first entry)  
 XX  
 DE Human secreted protein ax318\_3 cDNA.  
 XX  
 KW Secreted protein; human; ax318\_3; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 CDS 521..1114  
 FT /\*tag= a  
 FT sig\_peptide 563..601  
 FT /\*tag= b  
 FT mat\_peptide 602..1111  
 FT /\*tag= c  
 XX  
 PN WO9840404-A2.  
 XX  
 PD 17-SEP-1998.  
 XX  
 PF 09-MAR-1998; 98WO-US04601.  
 XX  
 PR 06-MAR-1998; 98US-0036321.  
 PR 11-MAR-1997; 97US-0815381.  
 XX  
 PA (GENY ) GENETICS INST INC.  
 XX  
 PI Agostino MJ, Jacobs K, LaVallie ER, McCoy JM, Merberg D;  
 PI Racie LA, Spaulding V, Treacy M;  
 XX  
 DR WPI: 1998-520802/44.  
 DR P-PSDB; AAW69423.  
 XX  
 PT New isolated polynucleotides and secreted proteins - useful as, e.g.  
 PT nutritional additives, immunostimulators, haematopoiesis regulators  
 PT and as diagnostic agents  
 XX  
 PS Claim 1; Page 66; 119pp; English.  
 XX  
 CC This full-length cDNA clone, designated ax318\_3, codes for a novel  
 CC human secreted protein (see AAW69423), i.e. ax318\_3 protein. It was  
 CC isolated from a human adult testis cDNA library using methods which  
 CC are selective for cDNAs encoding secreted proteins, or was  
 CC identified as encoding a secreted or transmembrane protein on the  
 CC basis of computer analysis of the encoding protein. Homology is  
 CC shown to some database sequences. The invention provides isolated  
 CC polynucleotides (see AAV58754-63) obtained from human adult testis,  
 CC brain, retina or placenta, or from foetal kidney or brain cDNA  
 CC libraries. These are all deposited as ATCC 98353. They encode  
 CC novel human secreted proteins (see AAW69423-33) that may have e.g.  
 CC nutritional activity, immune stimulating or suppressing activity,  
 CC haematopoiesis regulating activity, tissue growth activity,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity,

CC haemostatic and thrombolytic activity, receptor/ligand activity,  
 CC antiinflammatory activity, cadherin/tumour invasion suppressor  
 CC activity, tumour inhibition or other activities. They may also be  
 CC used for diagnostic purposes.  
 XX  
 SQ Sequence 1328 BP; 446 A; 307 C; 282 G; 293 T; 0 other;

Query Match 67.2%; Score 16.8; DB 19; Length 1328;  
 Best Local Similarity 90.0%; Pred. No. 3.2e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 tgctggatctgctgtttgaa 22  
 ||||| ||||| |||||  
 Db 567 tgctgtatctgctgtttgaa 586

RESULT 45  
 AAQ06001  
 ID AAQ06001 standard; DNA; 12284 BP.  
 XX  
 AC AAQ06001;  
 XX  
 DT 16-JAN-1991 (first entry)  
 XX  
 DE Sequence encoding protein characteristic of hog cholera virus (HCV).  
 XX  
 KW Vaccine; parvovirus; swine influenza virus.  
 XX  
 OS Hog cholera virus.  
 XX  
 FH Key Location/Qualifiers  
 CDS 364..12057  
 FT /\*tag= a  
 XX  
 PN EP389034-A.  
 XX  
 PD 26-SEP-1990.  
 XX  
 PF 12-MAR-1990; 90EP-0200573.  
 XX  
 PR 19-MAR-1989; 89EP-0104921.  
 PR 12-MAR-1990; 90EP-0200573.  
 XX  
 PA (ALKU ) AKZO NV.  
 XX  
 PI Meyers G, Rumenapf T, Thiel HJ;  
 XX  
 DR WPI: 1990-291979/39.  
 DR P-PSDB; AAR06996.  
 XX  
 PT New hog cholera virus vaccine and diagnostic - comprises nucleic  
 PT acid sequence of poly-peptide characteristic of hog cholera virus  
 XX  
 PS Claim 2; Fig 2; 29pp; English.  
 XX  
 CC Gene product may be used to provide a vaccine and Abs for diagnosis  
 CC of hog cholera viral infection in pigs.  
 XX  
 SQ Sequence 12284 BP; 3793 A; 2601 C; 3228 G; 2662 T; 0 other;

Query Match 67.2%; Score 16.8; DB 11; Length 12284;  
 Best Local Similarity 90.0%; Pred. No. 4.2e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 tggtatctgctgtttgaa 25  
 ||||| ||||| |||||  
 Db 2249 tggtatctgctgtttgaa 2268

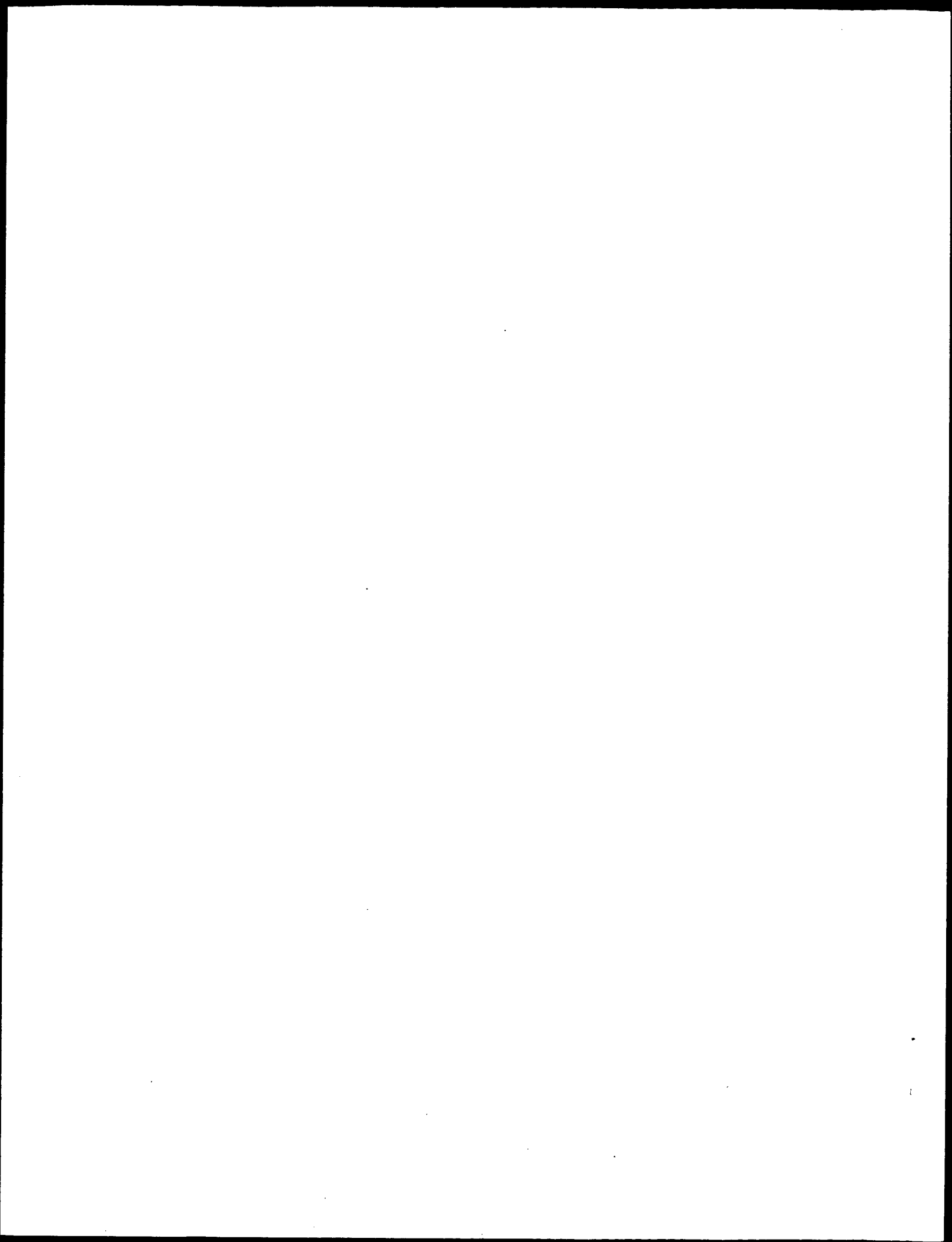
Search completed: October 9, 2001, 11:37:07

Wed Oct 10 07:44:31 2001

us-09-396-196f-3.std.rng

Page 29

Job time: 1742 sec







Query Match 100.0%; Score 25; DB 2; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 0.017;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctggtctgtctgtttgaagcg 25  
|||||  
Db 57 GTTGTGGATCTGCTGTTTGAAGCG 81

## RESULT 2

US-08-846-338-7  
; Sequence 7, Application US/08846338  
; Patent No. 5869719  
; GENERAL INFORMATION:  
; APPLICANT: Patton, David  
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5869719artis Corporation  
; STREET: 520 White Plains Road, P.O. Box 2005  
; CITY: Tarrytown  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10591  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30B  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/846.338  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8587  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1041 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1038  
; IDENTIFICATION METHOD: experimental  
; OTHER INFORMATION: /product= "biotin synthase"  
; OTHER INFORMATION: /evidence= EXPERIMENTAL

US-08-846-338-7  
Query Match 100.0%; Score 25; DB 2; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 0.017;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctggtctgtctgtttgaagcg 25  
|||||  
Db 57 GTTGTGGATCTGCTGTTTGAAGCG 81

## RESULT 3

US-08-411-768B-1  
; Sequence 1, Application US/08411768B  
; Patent No. 6083712  
; GENERAL INFORMATION:  
; APPLICANT: Olwen Birch  
; APPLICANT: Johann Brass  
; APPLICANT: Martin Fuhrmann  
; APPLICANT: Nicholas Shaw  
; TITLE OF INVENTION: Biotechnological Method  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect  
; SOFTWARE: Version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/411,768B  
; FILING DATE: 31-March-95  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: CH 3124/92  
; FILING DATE: 02-OCT-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: CH 2134/93  
; FILING DATE: 15-JUL-1993  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5872 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Escherichia coli  
; STRAIN: DSM498  
; IMMEDIATE SOURCE:  
; CLONE: pBO30A-15/9  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 117..1157  
; IDENTIFICATION METHOD: experimental  
; OTHER INFORMATION: /codon\_start= 117  
; OTHER INFORMATION: /product= "biotin synthase"  
; OTHER INFORMATION: /evidence= EXPERIMENTAL  
; OTHER INFORMATION: /gene= "biob"  
; OTHER INFORMATION: /number= 1  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 2295..3050  
; OTHER INFORMATION: /codon\_start= 2295  
; OTHER INFORMATION: /function= "involved in pimeloyl-CoA synthesis"  
; OTHER INFORMATION: /product= "protein"  
; OTHER INFORMATION: /gene= "bioc"  
; OTHER INFORMATION: /number= 3  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 3750..5039  
; IDENTIFICATION METHOD: experimental  
; OTHER INFORMATION: /codon\_start= 3750  
; OTHER INFORMATION: /EC\_number= 2.6.1.62  
; OTHER INFORMATION: /product= "DAPA synthase"  
; OTHER INFORMATION: /evidence= EXPERIMENTAL  
; OTHER INFORMATION: /gene= "biobA"  
; OTHER INFORMATION: /number= 5  
; OTHER INFORMATION: /standard\_name= "S-Adenosyl-L-methionine:8-amino-7-oxononanoate aminotransf.";  
; OTHER INFORMATION: /feature=

```
; LOCATION: 5098..5574
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 5098
; OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
; OTHER INFORMATION: /product= "protein"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ORF1"
; OTHER INFORMATION: /number= 6
; FEATURE:
; NAME/KEY: -10_signal
; LOCATION: 45..49
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "promoter ptac"
; FEATURE:
; NAME/KEY: -35_signal
; LOCATION: 23..28
; OTHER INFORMATION: /standard_name= "promoter ptac"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 105..119
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "bioB RBS no.9"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 2284..2297
; OTHER INFORMATION: /standard_name= "bioC RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 3742..3752
; OTHER INFORMATION: /standard_name= "bioA RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 5088..5100
; OTHER INFORMATION: /standard_name= "ORF1 RBS"
; FEATURE:
; NAME/KEY: terminator
; LOCATION: 5583..5644
; OTHER INFORMATION: /standard_name= "rho-independent
; OTHER INFORMATION: transcriptional terminator"
; FEATURE:
; NAME/KEY: stem_loop
; LOCATION: 5583..5605
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 1..96
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /function= "promoter ptac"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: WO 87/01391 B1
; FILING DATE: 26-AUG-1986
; PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1
```

```
Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 gttgctggtcgtctgtttgaagcg 25
|||||
Db 173 GTTGTGGATCTGCTGTTTGAGCG 197
```

```
RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
```

```
; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411,768B
; FILING DATE: 31-March-95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 3124/92
; FILING DATE: 02-OCT-1992
; APPLICATION DATA:
; APPLICATION NUMBER: CH 2134/93
; FILING DATE: 15-JUL-1993
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5872 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: DSM498
; IMMEDIATE SOURCE:
; CLONE: pHO30A15-9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1154..2308
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 1154
; OTHER INFORMATION: /EC_number= 2.3.1.47
; OTHER INFORMATION: /product= "KAPA synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "biof"
; OTHER INFORMATION: /number= 2
; OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3043..3753
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 3043
; OTHER INFORMATION: /EC_number= 6.3.3.3
; OTHER INFORMATION: /product= "DTB synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "bioD"
; OTHER INFORMATION: /number= 4
; OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 1141..1156
; OTHER INFORMATION: /standard_name= "bioF RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 3030..3045
; OTHER INFORMATION: /standard_name= "bioD RBS"
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: WO 87/01391 B1
```





```

? COUNTRY: U.S.A.
? ZIP: 20850
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/876,991
? FILING DATE: 16-JUN-1997
? CLASSIFICATION: 424
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/08/747,577
? FILING DATE:
? APPLICATION NUMBER: US/08/650,584
? FILING DATE:
? APPLICATION NUMBER: US/08/469,702
? FILING DATE:
? APPLICATION NUMBER: US/08/123,596
? FILING DATE:
? APPLICATION NUMBER: 07/797,554
? FILING DATE: 22-NOV-1991
? APPLICATION NUMBER: US 07/494,991
? FILING DATE: 16-MAR-1990
? CLASSIFICATION: 424
? ATTORNEY/AGENT INFORMATION:
? NAME: William M. Blackstone
? REGISTRATION NUMBER: 29,772
? REFERENCE/DOCKET NUMBER:
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (301) 258-5200
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 12284 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: cdna
? ORGANISM SOURCE:
? ORGANISM: Hog cholera virus
? STRAIN: Alfort
? CELL LINE: PK 15 and 38A1D
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 364..12060
? OTHER INFORMATION: /label= 435_kDA_protein
? FEATURE:
? NAME/KEY: primer_bind
? LOCATION: complement (2587..2619)
? OTHER INFORMATION: /label= primer_1
? FEATURE:
? NAME/KEY: primer_bind
? LOCATION: complement (2842..2880)
? OTHER INFORMATION: /label= primer_2
? FEATURE:
? NAME/KEY: variation
? LOCATION: replace(127, "c")
? FEATURE:
? NAME/KEY: variation
? LOCATION: replace(1522, "g")
? FEATURE:
? NAME/KEY: variation
? LOCATION: replace(10989, "t")
? US-08-876-991-1

```

Qy 1 gttgctggatctgctgttgaagcg 25  
||| ||| ||| ||| ||| ||| |||  
Db 16 GTGGCGGGAAACGCTGTTTGAAGCG 40

Query Match	67.23;	Score 16.8;	DB 2;	Length 12284;
Best Local Similarity	90.03;	Pred. No. 96;		
Matches	18;	Conservative	0;	Mismatches 2;
				Indels 0;
Qy	6	tggatctgctgtttgaagcg	25	
Db	2249	TTGCTCTGCTGTTTGAAGAG	2268	



```

RESULT 15
PCT-US93-11153-31
; Sequence 31, Application PC/TUS9311153
; GENERAL INFORMATION:
; APPLICANT: Godiska, Ronald
; APPLICANT: Gray, Patrick W.
; APPLICANT: Schweikart, Vicki L.
; TITLE OF INVENTION: Novel Seven Transmembrane Receptors
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11153
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/977,452
; FILING DATE: 17-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Noland, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31794
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: exon
; LOCATION: 7..80
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 94..1158
PCT-US93-11153-31

```

```

Query Match 66.4%; Score 16.6; DB 5; Length 1161;
Best Local Similarity 82.6%; Pred. No. 82;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgctggatctgctgtttgaagc 24
||| ||||| ||||| |||||
DB 316 TTGCTGATCTGCTGTTGTAGC 338

```

```

RESULT 16
US-08-287-001A-1/c
; Sequence 1, Application US/08287001A
; Patent No. 5622861
; GENERAL INFORMATION:
; APPLICANT: KAPLAN, GERARDO
; APPLICANT: FEINSTONE, STEPHEN M.
; TITLE OF INVENTION: HEPATITIS A VIRUS RECEPTOR AND METHODS
; FILING DATE:
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: Suite 1200, The Candler Bldg, 127 Peachtree
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/287,001A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwedyln D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2093 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 196..1551
; US-08-287-001A-1

```

```

Query Match 66.4%; Score 16.6; DB 1; Length 2093;
Best Local Similarity 82.6%; Pred. No. 90;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gttgctggatctgctgtttgaag 23
||| ||||| ||||| |||||
DB 1402 GTTGTGAATCTCCTTTTGAAG 1380

```

```

RESULT 17
PCT-US95-09941-1/c
; Sequence 1, Application PC/TUS9509941
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HEPATITIS A VIRUS RECEPTOR AND METHODS
; TITLE OF INVENTION: OF USE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: Suite 1200, The Candler Bldg, 127 Peachtree
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/09941
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/287,001
; FILING DATE: 5 AUG 1994

```



```
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwedolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414,621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2093 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 196..1551
PCT-US95-09941-1

Query Match      66.4%; Score 16.6; DB 5; Length 2093;
Best Local Similarity 82.6%; Pred. No. 90;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 gttgctgactgtctgtttgaag 23
||||| ||||| ||||| ||||| |||||
Db 1402 GTTGTGAATCTCTCTTTTGAAG 1380

RESULT 18
US-08-153-848-27
; Sequence 27, Application US/08153848
; Patent No. 5759804
; GENERAL INFORMATION:
; APPLICANT: Godiska, Ronald
; APPLICANT: Gray, Patrick W.
; APPLICANT: Schweikart, Vicki L.
; TITLE OF INVENTION: No. 5759804el Seven Transmembrane Receptors
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/153,848
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/977,452
; FILING DATE: 17-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5759804and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31794
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2254 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 593..1657
US-08-153-848-27

Query Match      66.4%; Score 16.6; DB 1; Length 2254;
Best Local Similarity 82.6%; Pred. No. 91;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ttgctgactgtctgtttgaagc 24
||||| ||||| ||||| ||||| |||||
Db 816 TTGCTGATCTGCTGTTGTAGC 838

RESULT 19
US-09-299-843A-27
; Sequence 27, Application US/09299843A
; Patent No. 6107475
; GENERAL INFORMATION:
; APPLICANT: Godiska, Ronald
; APPLICANT: Gray, Patrick W.
; APPLICANT: Schweikart, Vicki L.
; TITLE OF INVENTION: No. 6107475el Seven Transmembrane Receptors
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,843A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/088,337
; FILING DATE: 01-JUN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,848
; FILING DATE: 17-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/977,452
; FILING DATE: 17-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Jill E. Uhl
; REGISTRATION NUMBER: 43,213
; REFERENCE/DOCKET NUMBER: 27866/32059B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX:
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2254 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 593..1657
US-09-299-843A-27
```

Query Match 66.4%; Score 16.6; DB 3; Length 2254;  
Best Local Similarity 82.6%; Pred. No. 91;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgctggtatcgtctgttgaagc 24  
||| ||||| ||||| |||  
Db 816 TTGCTGATCTGCTGTTGTAGC 838

## RESULT 20

PCT-US93-11153-27  
; Sequence 27, Application PC/TUS9311153  
; GENERAL INFORMATION:  
; APPLICANT: Godiska, Ronald  
; APPLICANT: Gray, Patrick W.  
; APPLICANT: Schweikart, Vicki L.  
; TITLE OF INVENTION: Novel Seven Transmembrane Receptors  
; NUMBER OF SEQUENCES: 64  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
; ADDRESSEE: Bicknell  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/11153  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/977,452  
; FILING DATE: 17-NOV-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Noland, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 31794  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 474-6300  
; TELEFAX: (312) 474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2254 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 593..1657  
PCT-US93-11153-27

Query Match 66.4%; Score 16.6; DB 5; Length 2254;  
Best Local Similarity 82.6%; Pred. No. 91;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgctggtatcgtctgttgaagc 24  
||| ||||| ||||| |||  
Db 816 TTGCTGATCTGCTGTTGTAGC 838

## RESULT 21

US-09-299-843A-31  
; Sequence 31, Application US/09299843A  
; Patent No. 6107475  
; GENERAL INFORMATION:

; APPLICANT: Godiska, Ronald  
; APPLICANT: Gray, Patrick W.  
; APPLICANT: Schweikart, Vicki L.  
; TITLE OF INVENTION: No. 6107475el Seven Transmembrane Receptors  
; NUMBER OF SEQUENCES: 66  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
; ADDRESSEE: Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/299,843A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/088,337  
; FILING DATE: 01-JUN-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/153,848  
; FILING DATE: 17-NOV-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/977,452  
; FILING DATE: 17-NOV-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jill E. Ohl  
; REGISTRATION NUMBER: 43,213  
; REFERENCE/DOCKET NUMBER: 27866/32059B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 474-6300  
; TELEFAX: (312) 474-0448  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 31:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3119 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 7..80  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 94..1158  
US-09-299-843A-31

Query Match 66.4%; Score 16.6; DB 3; Length 3119;  
Best Local Similarity 82.6%; Pred. No. 95;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgctggtatcgtctgttgaagc 24  
||| ||||| ||||| |||  
Db 316 TTGCTGATCTGCTGTTGTAGC 338

## RESULT 22

US-09-058-489-14  
; Sequence 14, Application US/09058489  
; Patent No. 6103886  
; GENERAL INFORMATION:  
; APPLICANT: Whitehead Institute for Biomedical Research  
; APPLICANT: Lahn, Bruce  
; APPLICANT: Page, David  
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of

; TITLE OF INVENTION: the Y Chromosome  
; FILE REFERENCE: WHI97-08PA  
; CURRENT APPLICATION NUMBER: US/09/058,489  
; CURRENT FILING DATE: 1998-04-10  
; EARLIER APPLICATION NUMBER: 60/041,877  
; EARLIER FILING DATE: 1997-04-11  
; NUMBER OF SEQ ID NOS: 91  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 14  
; LENGTH: 3408  
; TYPE: DNA  
; ORGANISM: Human  
US-09-058-489-14

Query Match 66.4%; Score 16.6; DB 3; Length 3408;  
Best Local Similarity 82.6%; Pred. No. 97;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ttgtggatctgctgtttgaagc 24  
||| ||||| || |||||  
Db 2522 ttgtggatctctctgtttgaagc 2544

## RESULT 23

US-08-188-582-15/c  
; Sequence 15, Application US/08188582  
; Patent No. 5534410  
; GENERAL INFORMATION:  
; APPLICANT: Tjian, Robert  
; APPLICANT: Comai, Lucio  
; APPLICANT: Dynlacht, Brian D.  
; APPLICANT: Hoey, Timothy  
; APPLICANT: Ruppert, Siegfried  
; APPLICANT: Tanese, Naoko  
; APPLICANT: Wang, Edith  
; APPLICANT: Weinzierl, Robert O.J.  
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS, OF USE  
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFS AND METHODS OF USE  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/188,582  
; FILING DATE: 28-JAN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman, Richard A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3603 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; FEATURE:  
; NAME/KEY: CDS

; LOCATION: 1..2214  
US-08-188-582-15

Query Match 66.4%; Score 16.6; DB 1; Length 3603;  
Best Local Similarity 82.6%; Pred. No. 97;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gttgtggatctgctgtttgaag 23  
| ||||| ||||| |||||  
Db 997 GCTGCTGGCTCTGCTGGATGAAG 975

## RESULT 24

US-08-646-715-15/c  
; Sequence 15, Application US/08646715  
; Patent No. 5637686  
; GENERAL INFORMATION:  
; APPLICANT: Tjian, Robert  
; APPLICANT: Comai, Lucio  
; APPLICANT: Dynlacht, Brian D.  
; APPLICANT: Hoey, Timothy  
; APPLICANT: Ruppert, Siegfried  
; APPLICANT: Tanese, Naoko  
; APPLICANT: Wang, Edith  
; APPLICANT: Weinzierl, Robert O.J.  
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,  
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFS AND METHODS OF USE  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/646,715  
; FILING DATE: 09-MAY-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/188,582  
; FILING DATE: 28-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman, Richard A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3603 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..2214  
US-08-646-715-15

Query Match 66.4%; Score 16.6; DB 1; Length 3603;  
Best Local Similarity 82.6%; Pred. No. 97;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gttcgtgatctgctgtttgaag 23  
| | | | | | | | | | | | | | | |  
Db 997 GCTGCTGGCTCTGCTGATGAAG 975

RESULT 25  
US-09-058-489-13  
; Sequence 13, Application US/09058489  
; Patent No. 6103886  
; GENERAL INFORMATION:  
; APPLICANT: Whitehead Institute for Biomedical Research  
; APPLICANT: Lahn, Bruce  
; APPLICANT: Page, David  
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of  
; FILE REFERENCE: WHI97-08pA  
; CURRENT APPLICATION NUMBER: US/09/058,489  
; CURRENT FILING DATE: 1998-04-10  
; EARLIER FILING DATE: 1997-04-11  
; NUMBER OF SEQ ID NOS: 91  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 13  
; LENGTH: 5322  
; TYPE: DNA  
; ORGANISM: Human  
US-09-058-489-13

Query Match 66.4%; Score 16.6; DB 3; Length 5322;  
Best Local Similarity 82.6%; Pred. No. 1e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 19; Conservative 0

QY 2 ttgctggatctgctgtttgaagc 24  
| | | | | | | | | | | | | | | |  
Db 2522 ttgttgatcttctgttgaagc 2544

RESULT 26  
US-08-811-583-1/c  
; Sequence 1, Application US/08811583  
; Patent No. 6218142  
; GENERAL INFORMATION:  
; APPLICANT: Wassenecker, Michael  
; APPLICANT: Riedel, Leonhard  
; APPLICANT: Schiebel, Winfried  
; APPLICANT: Sanger, Heinz  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN  
; TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RDRP)  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & NEAVE  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/811,583  
; FILING DATE: 05-MAR-1997  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley, James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: WPG-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000

; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3731 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; ORIGINAL SOURCE:  
; ORGANISM: Tomato  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 194..3535  
US-08-811-583-1

Query Match 65.6%; Score 16.4; DB 4; Length 3731;  
Best Local Similarity 94.4%; Pred. No. 1.2e+02; Mismatches 1; Indels 0; Gaps 0;  
Matches 17; Conservative 0

QY 1 gttcgtgatctgctgtt 18  
| | | | | | | | | | | | | | | |  
Db 2422 GTTGTGGATCTGCTGTT 2405

RESULT 27  
US-08-649-046-1  
; Sequence 1, Application US/08649046  
; Patent No. 5912415  
; GENERAL INFORMATION:  
; APPLICANT: OLSZEWSKI, NEIL E.  
; APPLICANT: JACOBSEN, STEVEN E.  
; TITLE OF INVENTION: THE SPINDLY GENE, METHODS OF  
; TITLE OF INVENTION: IDENTIFICATION AND USE  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MUETING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.  
; STREET: 119 NORTH FOURTH STREET, SUITE 203  
; CITY: MINNEAPOLIS  
; STATE: MINNESOTA  
; COUNTRY: USA  
; ZIP: 55401  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/649,046  
; FILING DATE: 16-MAY-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MCCORMACK, MYRA H.  
; REGISTRATION NUMBER: 36,602  
; REFERENCE/DOCKET NUMBER: 110.00340101  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-305-1225  
; TELEFAX: 612-305-1228  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3476 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-649-046-1

Query Match 64.8%; Score 16.2; DB 2; Length 3476;  
Best Local Similarity 85.7%; Pred. No. 1.4e+02; Mismatches 3; Indels 0; Gaps 0;  
Matches 18; Conservative 0

QY 2 ttgtcgtgatctgctgtttgaa 22





```

RESULT 34
US-09-104-070-1
; Sequence 1, Application US/09104070
; Patent No. 6153741
; GENERAL INFORMATION:
; APPLICANT: Richards, Eric J.
; APPLICANT: Jeddeloh, Jeffrey A.
; TITLE OF INVENTION: DNA Methylation Gene from Plants
; FILE REFERENCE: Wash U CI-0014 US
; CURRENT APPLICATION NUMBER: US/09/104,070
; CURRENT FILING DATE: 1998-06-24
; EARLIER APPLICATION NUMBER: US 60/083,612
; EARLIER FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 5000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (782)...(1252)
; NAME/KEY: CDS
; LOCATION: (1354)...(1440)
; NAME/KEY: CDS
; LOCATION: (1549)...(1895)
; NAME/KEY: CDS
; LOCATION: (1976)...(2165)
; NAME/KEY: CDS
; LOCATION: (2251)...(2437)
; NAME/KEY: CDS
; LOCATION: (2559)...(2629)
; NAME/KEY: CDS

```

RESULT 35  
US-09-036-987A-1  
; Sequence 1, Application US/09036987A  
; Patent No. 6143526  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H.  
; APPLICANT: Broughton, Mary C.  
; APPLICANT: Crawford, Kathryn P.  
; APPLICANT: Madhuri, Krishnamurthy  
; APPLICANT: Merlo, Donald J.  
; APPLICANT: Treadway, Patti J.  
; APPLICANT: Turner, Jan R.  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; TITLE OF INVENTION: Production  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dow AgroSciences LLC Patent Department  
; STREET: 9330 Zionsville Road  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: USA  
; ZIP: 46268  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/036,987A  
; FILING DATE: 09-MAR-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stuart, Donald R

Query Match 63.28; Score 15.8; DB 2; Length 1141;



Best Local Similarity 89.5%; Pred. No. 1.8e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 tgcgtgacgtcgtgttga 21  
||||| ||||| ||||| |||||  
Db 495 TGCTTGATCAGCTGTTGA 513

## RESULT 38

US-08-867-087B-1  
; Sequence 1, Application US/08867087B  
; Patent No. 5990386  
; GENERAL INFORMATION:  
; APPLICANT: An, Gynheung  
; TITLE OF INVENTION: GENES CONTROLLING FLORAL DEVELOPMENT  
; TITLE OF INVENTION: AND APICAL DOMINANCE IN PLANTS  
; NUMBER OF SEQUENCES: 70  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klarquist Sparkman Campbell Leigh &  
; ADDRESSEE: Whinston, LLP  
; STREET: One World Trade Center  
; STREET: 121 S.W. Salmon Street  
; STREET: Suite 1600  
; CITY: Portland  
; STATE: Oregon  
; COUNTRY: United States of America  
; ZIP: 97204

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Disk, 3-1/2 inch  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: MS DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/867.087B  
; FILING DATE: June 2, 1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: U.S. 08/323,449  
; FILING DATE: October 14, 1994  
; APPLICATION NUMBER: U.S. 08/485,981  
; FILING DATE: June 7, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dow, Alan. E.  
; REGISTRATION NUMBER: 35,123  
; REFERENCE/DOCKET NUMBER: 4630-47071  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (503) 226-7391  
; TELEFAX: (503) 228-9446

INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1141 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double stranded  
; TOPOLOGY: linear  
; US-08-867-087B-1

Query Match 63.2%; Score 15.8; DB 2; Length 1141;  
Best Local Similarity 89.5%; Pred. No. 1.8e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 tgcgtgacgtcgtgttga 21  
||||| ||||| ||||| |||||  
Db 495 TGCTTGATCAGCTGTTGA 513

## RESULT 39

US-08-121-713D-57  
; Sequence 57, Application US/08121713D  
; Patent No. 5639856  
; GENERAL INFORMATION:  
; APPLICANT: Goodman, Corey S.  
; APPLICANT: Kolodkin, Alex L.

; APPLICANT: Matthes, David  
; APPLICANT: Bentley, David R.  
; APPLICANT: O'Connor, Timothy  
; TITLE OF INVENTION: The Semaphorin Gene Family  
; NUMBER OF SEQUENCES: 100  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 Bush Street, Suite 3200  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/121,713D  
; FILING DATE: 13-SEP-1993  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman, Richard A.  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: B94-002-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 343-4341  
; TELEFAX: (415) 343-4342  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 57:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2854 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 451..2640  
; US-08-121-713D-57

Query Match 63.2%; Score 15.8; DB 1; Length 2854;  
Best Local Similarity 89.5%; Pred. No. 2.1e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ttgctggatcgtcgtgttg 20  
||||| ||||| ||||| |||||  
Db 1382 TTGGTGGCTCTGCTGTTG 1400

## RESULT 40

US-08-835-268-57  
; Sequence 57, Application US/08835268  
; Patent No. 5807826  
; GENERAL INFORMATION:  
; APPLICANT: Goodman, Corey S.

; APPLICANT: Kolodkin, Alex L.  
; APPLICANT: Matthes, David  
; APPLICANT: Bentley, David R.  
; APPLICANT: O'Connor, Timothy  
; TITLE OF INVENTION: The Semaphorin Gene Family  
; NUMBER OF SEQUENCES: 100  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 Bush Street, Suite 3200  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

; REFERENCE/DOCKET NUMBER: B94-002-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415) 343-4342
; TELEX:
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2854 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 451..2640
; US-09-060-692-57
;
Query Match 63.2%; Score 15,8; DB 2; Length 2854;
Best Local Similarity 89.5%; Pred No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0
OY 2 ttgtcggatctgctgttgg 20
   ||| ||| ||| ||| ||| |||
Db 1382 TTGCTGGCTCTGCTGTTG 1400
;
RESULT 42
US-08-833-391-57
; Sequence 57, Application US/08833391
; Patent No. 6013781
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey S.
; APPLICANT: Kolodkin, Alex L.
; APPLICANT: Matthes, David
; APPLICANT: Bentley, David R.
; APPLICANT: O'Connor, Timothy
; TITLE OF INVENTION: The Semaphorin Gene Family
; NUMBER OF SEQUENCES: 100
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/833,391
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/121,713
; FILING DATE: 13-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B94-002-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415) 343-4342
; TELEX:
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2854 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna

```



```

; CITY: NEW YORK
; STATE: NY
; COUNTRY: USA
; ZIP: 10016-2391
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,874
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/377,440
; FILING DATE: 24-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: GOLDBERG, JULES E.
; REGISTRATION NUMBER: 24,408
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-986-4090
; TELEFAX: 212-818-9479
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28804 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FRAGMENT TYPE: N-terminal
; US-08-592-874-1

```

```

Query Match      63.2%; Score 15.8; DB 2; Length 28804;
Best Local Similarity 89.5%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 tgctggtctgctgtttga 21
   |||||
Db 14585 TGCTGGATCTGCTGTCGA 14567

```

```

Search completed: October 9, 2001, 11:39:37
Job time: 1892 sec

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 13:46:30 ; Search time 5930.9 Seconds  
(without alignments)  
39.846 Million cell updates/sec

Title: US-09-396-196f-3  
Perfect score: 25  
Sequence: 1 gttgctggtatctgctgtttgaagcg 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est25.\*  
26: gb\_est26.\*  
27: gb\_est27.\*  
28: gb\_est28.\*  
29: gb\_est29.\*  
30: gb\_est30.\*  
31: gb\_est31.\*  
32: gb\_est32.\*  
33: gb\_est33.\*  
34: gb\_est34.\*  
35: gb\_est35.\*  
36: gb\_est36.\*  
37: gb\_est37.\*  
38: gb\_est38.\*  
39: gb\_est39.\*  
40: gb\_est40.\*  
41: gb\_est41.\*  
42: gb\_est42.\*  
43: gb\_est43.\*  
44: gb\_est44.\*  
45: gb\_est45.\*  
46: gb\_est46.\*  
47: gb\_est47.\*

44: em\_esthum10.\*  
45: em\_esthum11.\*  
46: em\_esthum12.\*  
47: em\_esthum13.\*  
48: em\_esthum14.\*  
49: em\_esthum15.\*  
50: em\_esthum16.\*  
51: em\_esthum17.\*  
52: em\_esthum18.\*  
53: em\_esthum19.\*  
54: em\_esthum20.\*  
55: em\_esthum21.\*  
56: em\_esthum22.\*  
57: em\_esthum23.\*  
58: em\_esthum24.\*  
59: em\_esthum25.\*  
60: em\_esthum26.\*  
61: em\_esthum27.\*  
62: em\_esthum28.\*  
63: em\_estin1.\*  
64: em\_estin2.\*  
65: em\_estin3.\*  
66: em\_estin4.\*  
67: em\_estin5.\*  
68: em\_estom1.\*  
69: em\_estom2.\*  
70: em\_estov1.\*  
71: em\_estov2.\*  
72: em\_estpl1.\*  
73: em\_estpl2.\*  
74: em\_estpl3.\*  
75: em\_estpl4.\*  
76: em\_estpl5.\*  
77: em\_estpl6.\*  
78: em\_estpl7.\*  
79: em\_estpl8.\*  
80: em\_estpl9.\*  
81: em\_estpl10.\*  
82: em\_estro1.\*  
83: em\_estro2.\*  
84: em\_estro3.\*  
85: em\_estro4.\*  
86: em\_estro5.\*  
87: em\_estro6.\*  
88: em\_estro7.\*  
89: em\_estro8.\*  
90: em\_estro9.\*  
91: em\_estro10.\*  
92: em\_estro11.\*  
93: em\_estro12.\*  
94: em\_estro13.\*  
95: em\_estro14.\*  
96: em\_estro15.\*  
97: em\_estro16.\*  
98: em\_estro17.\*  
99: em\_estro18.\*  
100: em\_estro19.\*  
101: em\_estro20.\*  
102: gb\_est25.\*  
103: gb\_est26.\*  
104: gb\_est27.\*  
105: gb\_est28.\*  
106: gb\_est29.\*  
107: gb\_est30.\*  
108: gb\_est31.\*  
109: gb\_est32.\*  
110: gb\_est41.\*  
111: gb\_est42.\*  
112: gb\_est43.\*  
113: gb\_est44.\*  
114: gb\_est45.\*  
115: gb\_est46.\*  
116: gb\_est47.\*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
c 1	20.8	83.2	610	222	CNS051ED	AL316750 Tetraodon
c 2	20.2	80.8	465	120	AW766819	da71d10.x
c 3	20.2	80.8	618	118	AW636257	bl44f04.x
c 4	19.4	77.6	888	221	CNS04KH2	AL294815 Tetraodon
c 5	19.2	76.8	1101	219	CNS002W7	AL063545 Drosophil
c 6	18.8	75.2	412	175	BC233131	daa32f11.
c 7	18.8	75.2	412	175	BC233131	daa32f11.
c 8	18.6	74.4	290	137	BE586432	WHE504.D0
c 9	18.6	74.4	290	117	AW501847	UI-HF-BR0
c 10	18.6	74.4	300	117	AW503005	UI-HF-BR0
c 11	18.6	74.4	662	155	BG593982	EST492660
c 12	18.4	73.6	594	222	FR0016852	Mus muscu
c 13	18.4	73.6	849	220	CNS02BSU	AL124355 Fugu rubr
c 14	18.4	73.6	859	220	CNS02BSU	AL194151 Tetraodon
c 15	18.2	72.8	85	22	CNS03713	AL230736 Tetraodon
c 16	18.2	72.8	85	22	AI605313	AL605313 vn87h03.x
c 17	18.2	72.8	101	9	AA624863	AA624863 vn87h03.x
c 18	18.2	72.8	106	23	AI642276	AI642276 vn87h03.y
c 19	18.2	72.8	293	240	AZ214122	Sheared D
c 20	18.2	72.8	303	143	BE990318	UI-M-BX1-
c 21	18.2	72.8	345	172	BG008569	PM4-GN030
c 22	18.2	72.8	422	243	AZ412964	LM0186F13
c 23	18.2	72.8	437	20	AI449930	AI449930 mr81a05.x
c 24	18.2	72.8	437	229	AQ522448	HS-5237.A
c 25	18.2	72.8	464	258	TA372A07Q	AL496271 T. brucei
c 26	18.2	72.8	495	8	AA527258	ng35h02.S
c 27	18.2	72.8	499	119	AW702183	AW702183 TgSiz273
c 28	18.2	72.8	507	240	AZ244099	RPCI-23-3
c 29	18.2	72.8	517	138	BE656252	UI-M-BH0-
c 30	18.2	72.8	521	167	BE443324	WHE1112.E
c 31	18.2	72.8	532	258	TA375E11P	AL495591 T. brucei
c 32	18.2	72.8	541	236	AZ003494	RPCI-23-3
c 33	18.2	72.8	559	258	TA162C10P	AL472440 T. brucei
c 34	18.2	72.8	596	235	AQ936593	HSJ41-840
c 35	18.2	72.8	621	235	AQ941180	Sheared D
c 36	18.2	72.8	628	112	AI160263	EST290121
c 37	18.2	72.8	650	258	TA69C05P	AL457556 T. brucei
c 38	18.2	72.8	685	18	AI325366	mi29f01.Y
c 39	18.2	72.8	938	222	CNS0564D	AL322870 Tetraodon
c 40	18.2	72.8	1839	192	AK006021	Mus muscu
c 41	18.2	72.0	403	163	BE095496	UI-R-B00-
c 42	17.8	71.2	487	148	BF428015	daa04g08.
c 43	17.8	71.2	240	107	FR0044714	AL132206 Fugu rubr
c 44	17.8	71.2	295	148	BF440174	AU074974 AU074974
c 45	17.8	71.2	296	128	BB190815	BB190815 BB190815

## ALIGNMENTS

RESULT	1	CNS051ED	610 bp	DNA	GSS	26-JUL-2000
LOCUS	AL316750	AL316750				
DEFINITION	CNS051ED	Tetraodon nigroviridis genome survey sequence T7 end of clone				
ACCESSION	AL316750	AL316750				
VERSION	1	GI:9549634				
KEYWORDS	GSS: genome survey sequence.	GSS: genome survey sequence.				
SOURCE	Tetraodon nigroviridis.	Tetraodon nigroviridis.				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.				
REFERENCE	1 (bases 1 to 610)	1 (bases 1 to 610)				
AUTHORS	Roest Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,	Roest Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,				

Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.  
 Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence  
 Nat. Genet. 25 (2), 235-238 (2000)  
 20296633  
 2 (bases 1 to 610)  
 Crollius, H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C., Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.  
 Characterization and repeat analysis of the compact genome of the freshwater pufferfish tetraodon nigroviridis  
 Genome Res. 10 (7), 939-949 (2000)  
 20359837  
 3 (bases 1 to 610)  
 Genoscope.  
 Direct Submission  
 Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
 This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at  
 http://www.genoscope.cns.fr/tetraodon.  
 Location/Qualifiers  
 1. .610  
 /organism="Tetraodon nigroviridis"  
 /db\_xref="taxon:99883"  
 /clone="051011"  
 /note="Genoscope sequence ID : C0AA051AH06C1-end : T7"  
 BASE COUNT 124 a 159 c 162 g 139 t 26 others  
 ORIGIN  
 Query Match 83.2%; Score 20.8; DB 222; Length 610;  
 Best Local Similarity 91.7%; Pred. No. 73;  
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 ttgtgcatctgctgtttgaagcg 25  
 Db 338 TGGCTGGATCTGCAGTTGAAGCG 315  
 RESULT 2  
 LOCUS AW766819/c 465 bp mRNA EST 16-FEB-2001  
 DEFINITION IM7d1d10.x1 Harland stage 19-23 Xenopus laevis cDNA clone  
 AW766819  
 ACCESSION AW766819.1 GI:7698812  
 VERSION AW766819  
 KEYWORDS EST.  
 SOURCE African clawed frog.  
 ORGANISM Xenopus laevis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae; Xenopodinae; Xenopus.  
 1 (bases 1 to 465)  
 Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D., Martin, J., Wylie, I., Underwood, K., Theising, B., Bowers, Y., Person, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R., Waterston, R. and Wilson, R.  
 WashU Xenopus EST project, 1999  
 Unpublished (1999)  
 Contact: Sandy Clifton, Ph.D.  
 WashU Xenopus EST project, 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 Library constructed by R. Harland, PhD (University of California, Berkeley)  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: Xenopus clones from this library are available

through the I.M.A.G.E. Consortium/LLNL at:  
 image.llnl.gov/image/html/iresources.shtml  
 Seq primer: -40up from G1bco  
 High quality sequence stop: 462.

#### FEATURES

source  
 1. .465  
 /location/Qualifiers  
 /organism="Xenopus laevis"  
 /db\_xref="taxon:8355"  
 /clone="IMAGE:3200371"  
 /clone\_lib="Harland stage 19-23"  
 /tissue\_type="neurala"  
 /dev\_stage="stage 19-23"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Vector: PCS107 (custom); Site 1: NotI; Site 2: SalI  
 ; cDNA made by oligo-dT priming. Library constructed by  
 Dr. Francesca Mariani in the laboratory of R. Harland,  
 Ph.D. (University of California, Berkeley). References:  
 XRF-2 is a transcriptional repressor that converts  
 ectoderm into neural tissue. Mariani, F.V. Harland, R.M.,  
 Development. 1998 Dec;125(24):5019-31. PMID: 9811586; UI:  
 99030283; Use of large-scale expression cloning screens in  
 the xenopus laevis tadpole to identify gene function.  
 Grammer TC, Liu KJ, Mariani FV, Harland RM, Dev Biol.  
 2000 Dec 15;228(2):197-210. PMID: 11112324; UI: 20564075;  
 Note: This is a Xenopus Gene Collection (XGC) library."

#### BASE COUNT

141 a 93 c 76 g 155 t

Query Match 80.8%; Score 20.2; DB 120; Length 465;

Best Local Similarity 88.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gttgtggtatcgtgtttgaagcg 25

Db 28 GTTGCTGAATCTCGGTTTGAACG 4

#### RESULT 3

LOCUS AW636257 618 bp mRNA EST 03-APR-2000  
 DEFINITION b144f04.w1 Blackshear/Soares normalized Xenopus egg library Xenopus  
 laevis cDNA clone PBX0044F04 5', mRNA sequence.  
 ACCESSION AW636257  
 VERSION AW636257.1 GI:7393338  
 KEYWORDS EST  
 SOURCE African clawed frog.  
 ORGANISM Xenopus laevis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
 Xenopodinae; Xenopus.  
 REFERENCE 1 (bases 1 to 618)  
 AUTHORS Blackshear, P.J., Lai, W.S., Thorn, J.M., Kennington, E.A., Staffa, N.G.,  
 Jr., Moore, D.T., Bouffard, G.G., Beckstrom-Sternberg, S.M., Touchman  
 , J.W., Bonaldo, M.F. and Soares, M.B.  
 TITLE The NIEHS Xenopus Maternal EST Project  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Perry J. Blackshear  
 Office of Clinical Research and Laboratory of Signal Transduction  
 National Institute of Environmental Health Sciences  
 A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,  
 USA  
 Tel: 919 541-4899  
 Fax: 919 541-4571  
 Email: black009@niehs.nih.gov  
 Clone is available through Research Genetics, Inc., 2130 Memorial  
 Parkway, Huntsville, AL 35901  
 phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email  
 cdna@resgen.com  
 DNA Sequencing and analyses performed by National Institutes of  
 Health Intramural Sequencing Center (NISC).  
 PCR Primers  
 FORWARD: TGTAAACGACGCGCCAGT

BACKWARD: CAGCAACACGCTATGACC  
 Plate: 0044 row: F column: 04  
 Seq primer: T7 primer.

#### FEATURES

source  
 1. .618  
 /location/Qualifiers  
 /organism="Xenopus laevis"  
 /db\_xref="taxon:8355"  
 /clone="PBX0044F04"  
 /clone\_lib="Blackshear/Soares normalized Xenopus egg  
 library"  
 /sex="female"  
 /tissue\_type="unfertilized egg"  
 /cell\_type="unfertilized egg"  
 /dev\_stage="unfertilized egg"  
 /lab\_host="DH10B"

/note="Vector: pT7T3-Pac; Site 1: EcoRI; Site 2: NotI;  
 PolyA-selected mRNA was prepared from unfertilized Xenopus  
 laevis eggs. The library was constructed in the vector  
 pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and  
 Soares, M.B. 'Normalization and subtraction: two  
 approaches to facilitate gene discovery', Genome Research  
 6:791-806, 1996. The first strand synthesis used a  
 NotI-dt18 primer; double stranded cDNAs were ligated to  
 EcoRI adapters, digested with NotI, and directionally  
 cloned into the NotI and EcoRI-digested pT7T3-Pac vector.  
 The library contained approximately 7.2 X 10<sup>5</sup>  
 recombinants, with average insert sizes of 1-1.5 kb."

BASE COUNT 176 a 106 c 107 g 229 t

Query Match 80.8%; Score 20.2; DB 118; Length 618;

Best Local Similarity 88.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gttgtggtatcgtgtttgaagcg 25

Db 325 GTTGCTGAATCTCGGTTTGAACG 349

#### RESULT 4

LOCUS CNS04KH2 888 bp DNA GSS 21-MAY-2000  
 DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone  
 116D03 of library G from Tetraodon nigroviridis, genomic survey  
 sequence.  
 ACCESSION AL294815.1 GI:8033395  
 VERSION AL294815  
 KEYWORDS GSS; genome survey sequence.  
 SOURCE Tetraodon nigroviridis.  
 ORGANISM Tetraodon nigroviridis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 Tetraodontidae; Tetraodon.  
 REFERENCE 1 (bases 1 to 888)  
 AUTHORS Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizes, C.,  
 Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and  
 Weissenbach, J.  
 TITLE Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetraodon nigroviridis  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 888)  
 AUTHORS Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,  
 Bernot, A., Fizes, C., Wincker, P., Brottier, P., Quetier, F.,  
 Saurin, W. and Weissenbach, J.  
 TITLE Human gene number estimate provided by genome wide analysis using  
 Tetraodon nigroviridis DNA sequence  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 888)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases



COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

## FEATURES

source  
1..888  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="116D03"  
/clone\_lib="G"  
/note="Genoscope sequence ID : C0RG116C802LP1-end : T7"  
BASE COUNT 129 a 276 c 276 g 187 t 20 others  
ORIGIN

Query Match 77.6%; Score 19.4; DB 221; Length 888;  
Best Local Similarity 95.2%; Pred. No. 3.1e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3 tgcgtgactcgtctgtttgaag 23  
|||||  
Db 331 TGCCTGATCTGCTGTGAG 351  
|||||

## RESULT 5

CNS002W7 1101 bp DNA GSS 03-JUN-1999  
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #  
DEFINITION BACH07D12 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL063545  
VERSION AL063545.1 GI:4941301  
KEYWORDS GSS:  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
Bp 191 91006 EVRY cedex - FRANCE (E-mail : [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr))  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammos in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

Location/Qualifiers  
1..1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"  
/clone="BACH07D12"  
/note="end : TET3"  
BASE COUNT 333 a 235 c 227 g 289 t 17 others  
ORIGIN

## FEATURES

source  
1..1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"  
/clone="BACH07D12"  
/note="end : TET3"  
BASE COUNT 333 a 235 c 227 g 289 t 17 others  
ORIGIN

Query Match 76.8%; Score 19.2; DB 219; Length 1101;  
Best Local Similarity 87.5%; Pred. No. 3.9e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgctgactcgtctgtttgaagc 25  
|||||  
Db 920 TTGCGGATGCTCTGTGAAGC 943  
|||||

RESULT 6  
BG233131 412 bp mRNA EST 12-FEB-2001  
LOCUS dsa32f11.y1 NICHHD XGC Lul Xenopus laevis cDNA clone IMAGE:4057893  
DEFINITION 5' similar to SW:AN3\_XENLA P24346 PUTATIVE ATP-DEPENDENT RNA  
HELICASE AN3.; mRNA sequence.  
ACCESSION BG233131  
VERSION BG233131.1 GI:12746978  
KEYWORDS EST.  
SOURCE African clawed frog.  
ORGANISM Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Xenopus.  
1 (bases 1 to 412)  
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps@email.nih.gov](mailto:cgaps@email.nih.gov)  
Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LLNL at: [info@image.llnl.gov](mailto:info@image.llnl.gov)  
Seq primer: -40RP from Gibco  
High quality sequence stop: 403.

Location/Qualifiers  
1..412  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/clone="IMAGE:4057893"  
/clone\_lib="NICHHD XGC Lul"  
/dev\_stage="adult"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pCMV-SPORT6; Site: 1; NotI; Site2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.6 Kb. Constructed by Life Technologies. Note: This is a Xenopus Gene Collection (XGC) library."  
BASE COUNT 115 a 79 c 121 g 97 t  
ORIGIN

Query Match 75.2%; Score 18.8; DB 175; Length 412;  
Best Local Similarity 90.9%; Pred. No. 5.1e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 tgcgtgactcgtctgtttgaagc 24  
|||||  
Db 52 TGCTGGATCTGCTGTGAAGC 73  
|||||

RESULT 7  
BE586432 498 bp mRNA EST 17-AUG-2000  
LOCUS WHE504\_D03\_H06ZR Secale cereale aluminum-stressed root tip cDNA  
DEFINITION library Secale cereale cDNA clone WHE504\_D03\_H06, mRNA sequence.  
ACCESSION BE586432  
VERSION BE586432.1 GI:9839464  
KEYWORDS EST.  
SOURCE rye.  
ORGANISM Secale cereale  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 498)  
Anderson, O.D., Butler, E., Chao, S., Gustafson, J.P., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L. and Tong, J.C.  
The structure and function of the expressed portion of the wheat genomes - Aluminum-stressed root tip cDNA library from rye (Secale cereale)  
Unpublished (2000)  
Contact: Olin Anderson  
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818  
Email: oanderson@pw.usda.gov  
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20  
Seq primer: M13 reversed primer.

FEATURES  
source  
Location/Qualifiers  
1..498  
/organism="Secale cereale"  
/cultivar="Blanco"  
/db\_xref="taxon:4550"  
/clone="WHE504\_D03\_H06"  
/clone\_lib="Secale cereale aluminum-stressed root tip cDNA library"  
/tissue\_type="Root tip"  
/dev\_stage="Seedling"  
/lab\_host="E. coli DH12S"  
/note="Vector: pSPORT1; Site\_1: SalI; Site\_2: NotI; Seeds were germinated aseptically on filter paper and transferred to a hydroponic growth system in a growth chamber when the primary root was 1 cm in length. After a 2-day establishment period, seedlings were subjected to a 5 ppm aluminum stress prior to tissue harvest. Plants were grown in an environmental chamber. The tissue, total RNA, and poly(A) RNA were prepared, and a cDNA library was made (Butler and Gustafson) at University of Missouri, Columbia. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT  
ORIGIN

125 a 96 c 134 g 142 t 1 others

Query Match 75.2%; Score 18.8; DB 137; Length 498;  
Best Local Similarity 90.9%; Pred. No. 5.2e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ttgctggatctgctgtttgaag 23  
|||||  
DB 37 TTGCTGGATCTCCTGTGTGAAG 16  
|||||

RESULT 8  
AW501847/c  
LOCUS  
DEFINITION  
UI-HF-BR0p-ajo-e-04-0-UI.r1 NIH\_MGC\_52 Homo sapiens cDNA clone  
IMAGE:3075174 5', mRNA sequence.  
ACCESSION  
AW501847  
VERSION  
AW501847.1 GI:7115764  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 290)  
/db\_xref="taxon:9606"  
/clone="IMAGE:3076393"  
/clone\_lib="NIH\_MGC\_52"  
/tissue\_type="lymph"  
/cell\_type="germinal center B cells"  
/cell\_line="MGC85"

Query Match 75.2%; Score 18.8; DB 137; Length 498;  
Best Local Similarity 90.9%; Pred. No. 5.2e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ttgctggatctgctgtttgaag 23  
|||||  
DB 37 TTGCTGGATCTCCTGTGTGAAG 16  
|||||

RESULT 8  
AW501847/c  
LOCUS  
DEFINITION  
UI-HF-BR0p-ajo-e-04-0-UI.r1 NIH\_MGC\_52 Homo sapiens cDNA clone  
IMAGE:3075174 5', mRNA sequence.  
ACCESSION  
AW501847  
VERSION  
AW501847.1 GI:7115764  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 290)  
/db\_xref="taxon:9606"  
/clone="IMAGE:3076393"  
/clone\_lib="NIH\_MGC\_52"  
/tissue\_type="lymph"  
/cell\_type="germinal center B cells"  
/cell\_line="MGC85"

FEATURES  
source  
Location/Qualifiers  
1..290  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3075174"  
/clone\_lib="NIH\_MGC\_52"  
/tissue\_type="lymph"  
/cell\_type="germinal center B cells"  
/cell\_line="MGC85"  
/lab\_host="DH10B (LT1)"  
/note="Vector: pT7T3-Pac; Site\_1: NotI; Site\_2: Eco RI; Constructed from size fractionated cytoplasmic mRNA (7.4-9.5kb). Directionally cloned. Cells provided by Louis M. Staudt, Ph.D. Library preparation by Maria de Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT  
ORIGIN

60 a 108 c 99 g 23 t

Query Match 74.4%; Score 18.6; DB 117; Length 290;  
Best Local Similarity 84.0%; Pred. No. 5.9e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gtgtggtggtctgctgtttgaagc 25  
|||||  
DB 236 GTGCTGCATCTGCTGTCGAAGCG 212  
|||||

RESULT 9  
AW503005/c  
LOCUS  
DEFINITION  
UI-HF-BR0p-ajo-e-01-0-UI.r1 NIH\_MGC\_52 Homo sapiens cDNA clone  
IMAGE:3076393 5', mRNA sequence.  
ACCESSION  
AW503005  
VERSION  
AW503005.1 GI:7118004  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 300)  
/db\_xref="taxon:9606"  
/clone="IMAGE:3076393"  
/clone\_lib="NIH\_MGC\_52"  
/tissue\_type="lymph"  
/cell\_type="germinal center B cells"  
/cell\_line="MGC85"

BASE COUNT  
ORIGIN

60 a 108 c 99 g 23 t

Query Match 74.4%; Score 18.6; DB 117; Length 290;  
Best Local Similarity 84.0%; Pred. No. 5.9e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gtgtggtggtctgctgtttgaagc 25  
|||||  
DB 236 GTGCTGCATCTGCTGTCGAAGCG 212  
|||||

RESULT 9  
AW503005/c  
LOCUS  
DEFINITION  
UI-HF-BR0p-ajo-e-01-0-UI.r1 NIH\_MGC\_52 Homo sapiens cDNA clone  
IMAGE:3076393 5', mRNA sequence.  
ACCESSION  
AW503005  
VERSION  
AW503005.1 GI:7118004  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 300)  
/db\_xref="taxon:9606"  
/clone="IMAGE:3076393"  
/clone\_lib="NIH\_MGC\_52"  
/tissue\_type="lymph"  
/cell\_type="germinal center B cells"  
/cell\_line="MGC85"

Eco RI site shown at the beginning of the sequence.  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: M.B. Soares Lab  
cDNA Library Arrayed by: M.B. Soares Lab  
DNA Sequencing by: M.B. Soares Lab  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html  
Seq primer: M13 Forward.

Eco RI site shown at the beginning of the sequence.  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: M.B. Soares Lab  
cDNA Library Arrayed by: M.B. Soares Lab  
DNA Sequencing by: M.B. Soares Lab  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html  
Seq primer: M13 Forward.



GAGAGAGATTCGAGTCTAATAATTAATCCCGCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda PLC I. Cloning sites, 5', end: SalI; 3', end: BamHI. Host: DH10B.

## FEATURES

## source

Location/Qualifiers

1..1329  
/organism="Mus musculus"  
/strain="C57BL/6j"  
/db\_xref="taxon:10090"  
/db\_xref="MGD:MGI:1908486"  
/db\_xref="MGD:MGI:1922422"  
/clone="4930528603"  
/sex="male"  
/tissue\_type="testis"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
125..1216  
/note="putative"  
/codon\_start=1  
/protein\_id="BAB30036.1"  
/db\_xref="GI:12854458"  
/translation="MEYNSKDAEETEDEERGLKEARGEQKEEVATSTETDDQED  
LPVLVPVNTREERLINLAETPAFLCLHELHSGKLPGLPTGRLMAELKAKYTLHDTVVS  
TQSEVOLLNAKRFTEIQOQQVCLQQAEDFPNVFTVECKLREQLLKYNQVYTAQ  
ERXNIOVRLTSLTEESIIILKEFKIPKGEIEKTKTEELKESTEELRKEVIOQLLEI  
KNLRDVLVKOKLVREKLEELMEYQVGLKDDVHHQSPVQITREIKMKTRKKVE  
TEKNIVLESSELKLSLAKLENKVNFLAERDDIMKEVEGKRTFLEVKEREYQGLL  
KLEUTRENEASSLAERLLFLHILSPKF"  
1310..1315  
/note="putative"  
1329  
/note="putative"  
1329

## CDS

polyA\_signal  
polyA\_site  
BASE COUNT 481 a 243 c 345 g 260 t  
ORIGIN  
Query Match 74.4%; Score 18.6; DB 192; Length 1329;  
Best Local Similarity 84.0%; Pred. No. 7.2e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gttggtgattctgctgtttgaagc 25  
Db 494 GCTGCTGATCTGCTCCGTAAGCG 470

RESULT 12  
FR0036852  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

FR0036852 594 bp DNA GSS 22-OCT-1999  
Fugu rubripes GSS sequence, clone 039N10aE9, genomic survey  
sequence.  
AL124355  
AL124355.1 GI:6105970  
GSS: genome survey sequence.  
Takifugu rubripes  
Takifugu rubripes  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Takifugu.  
1 (bases 1 to 594)  
Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,  
Umrania,Y., Williams,G., and Brenner,S.  
Direct Submission  
Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource  
Centre, Hinxton, Cambridge, CB10 1SB. UK Email:  
biohelp@hmp.mrc.ac.uk  
Vector: pBluescript II KS  
V\_Type: phagemid  
PRIMER: KS  
DESCR:  
One pass dye-terminator sequencing of cosmid cloned genomic  
sequence.

## FEATURES

Location/Qualifiers

1..594  
/organism="Takifugu rubripes"  
/db\_xref="taxon:31033"  
/clone\_lib="cosmid 039N10"  
/clone="039N10aE9"  
BASE COUNT 138 a 135 c 144 g 157 t 20 others  
ORIGIN  
Query Match 73.6%; Score 18.4; DB 222; Length 594;  
Best Local Similarity 90.5%; Pred. No. 7.9e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 tctggtgattctgctgtttgaag 23  
Db 469 TNCIGAACTGCTGTTGAAG 489

RESULT 13  
CNS02ESU/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

CNS02ESU 849 bp DNA GSS 13-MAY-2000  
Tetraodon nigroviridis genome survey sequence pUC-Ori end of clone  
131G13 of library G from Tetraodon nigroviridis, genomic survey  
sequence.  
AL194151  
AL194151.1 GI:7832257  
GSS: genome survey sequence.  
Tetraodon nigroviridis.  
Tetraodon nigroviridis.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodon.  
1 (bases 1 to 849)  
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,  
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
Weissenbach,J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
Unpublished  
2 (bases 1 to 849)  
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.  
Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence  
Unpublished  
3 (bases 1 to 849)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetraodon.  
Location/Qualifiers  
1..849  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="131G13"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COAG131AD07SP1-end :  
pUC-Ori"

BASE COUNT 233 a 211 c 193 g 208 t 4 others  
ORIGIN  
Query Match 73.6%; Score 18.4; DB 220; Length 849;  
Best Local Similarity 95.0%; Pred. No. 8.3e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 ctggtgattctgctgtttgaagc 24  
||||| |||||||

Db 167 CTGAGCTGCTGTTGAAGC 148

RESULT 14  
CNS03713/c  
LOCUS  
DEFINITION  
Tetraodon nigroviridis genome survey sequence T7 end of clone  
001G13 of library G from Tetraodon nigroviridis, genomic survey  
sequence.

ACCESSION  
AL230736  
VERSION  
AL230736.1 GI:7889731  
KEYWORDS  
GSS: genome survey sequence.  
SOURCE  
Tetraodon nigroviridis.  
ORGANISM  
Tetraodon nigroviridis

REFERENCE  
1 (bases 1 to 859)  
Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,  
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
Weissenbach,J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis

TITLE  
Unpublished

JOURNAL  
2 (bases 1 to 859)  
Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.  
Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence

TITLE  
Unpublished

JOURNAL  
3 (bases 1 to 859)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES  
source  
1..859  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="001G13"  
/clone\_lib="G"  
/note="Genoscope sequence ID : C0BG001AD07LP1-end : T7"

BASE COUNT 254 a 199 c 178 g 228 t

ORIGIN

Query Match 73.6%; Score 18.4; DB 220; Length 859;  
Best Local Similarity 95.0%; Pred. No. 8.3e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 ctggatctgctgtttgaagc 24  
||||| ||||||| |||||

Db 242 CTGGAGCTGCTGTTGAAGC 223

RESULT 15  
AI605313/c  
LOCUS  
DEFINITION  
vnr87h03.x1 Stratagene mouse heart (#937316) Mus musculus cDNA clone  
IMAGE:1038965 3', mRNA sequence.

ACCESSION  
AI605313  
VERSION  
AI605313.1 GI:4614480  
KEYWORDS  
EST.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus

REFERENCE  
1 (bases 1 to 85)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

AUTHORS  
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person  
B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter  
E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,  
Waterston,R. and Wilson,R.  
The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)  
Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LLNL : contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:578989  
This clone was previously sequenced on the 5' end only, this new  
data is from the 3' end.

FEATURES  
source  
1..85  
/organism="Mus musculus"  
/strain="NIH/Swiss"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1038965"  
/clone\_lib="Stratagene mouse heart (#937316)"  
/sex="pooled"  
/tissue\_type="heart"  
/dev\_stage="13 day embryos"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: heart; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts.  
Average insert size: 1.0 kb; Uni-ZAP XR Vector: ~5'  
adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor  
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'.

BASE COUNT 24 a 21 c 21 g 19 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 22; Length 85;  
Best Local Similarity 87.0%; Pred. No. 7.6e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ttgctggatctgctgtttgaagc 24  
||||| ||||||| |||||

Db 31 TTGCTGGATCTGCCGTTGGAGC 9

RESULT 16  
AA624863  
LOCUS  
DEFINITION  
vnr87h03.r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone  
IMAGE:1038965 5', mRNA sequence.

ACCESSION  
AA624863  
VERSION  
AA624863.1 GI:2528739  
KEYWORDS  
EST.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus

REFERENCE  
1 (bases 1 to 101)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

AUTHORS  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800

Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:578989  
 Seq primer: -28ml3 rev1 ET from Amersham.

#### FEATURES

Location/Qualifiers  
 1..101  
 /organism="Mus musculus"  
 /strain="NIH/Swiss"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1038965"  
 /clone\_lib="Stratagene mouse heart (#937316)"  
 /sex="pooled"  
 /tissue\_type="heart"  
 /dev\_stage="13 day embryos"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: heart; Vector: pBluescript SK-; Site:1:  
 EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:  
 Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts.  
 Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'  
 adaptor sequence: 5' GAATTCGCACGAG 3' -3' adaptor  
 sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'

BASE COUNT 24 a 23 c 25 g 29 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 9; Length 101;  
 Best Local Similarity 87.0%; Pred. No. 7.8e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgctggtatctgtgttgaagc 24  
 |||||

DB 79 TTGCTGGATCTGCCGTTGGAGC 101

#### RESULT 17

LOCUS AI642276 106 bp mRNA 29-APR-1999  
 DEFINITION vn87h03.y1 Stratagene mouse heart (#937316) Mus musculus cDNA clone  
 IMAGE:1038965 5', mRNA sequence.

ACCESSION AI642276  
 VERSION AI642276.1 GI:4720751  
 KEYWORDS EST  
 SOURCE house mouse.

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 106)  
 AUTHORS Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person,  
 B., Swaller, F., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter,  
 E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
 Waterston, R. and Willson, R.

TITLE The WashU-NCI Mouse EST Project 1999  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:578989

This read is a RESEQUENCE of a previously sequenced mouse clone  
 This read has been verified (found to hit its original self in the  
 correct orientation)

Seq primer: -40RP from Gibco.  
 Location/Qualifiers  
 1..106  
 /organism="Mus musculus"  
 /strain="NIH/Swiss"

#### FEATURES

Location/Qualifiers  
 1..293  
 /organism="Trypanosoma brucei"  
 /strain="TREU927/4 Gutat 10.1"  
 /db\_xref="taxon:5691"  
 /clone="Sheared DNA-67G1"  
 /clone\_lib="Sheared DNA"  
 /note="Vector: pUC18; Site:1: SmaI; Constructed at The  
 Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from a cloned population of  
 Trypanosoma brucei (TREU927/4 Gutat 10.1) was mechanically  
 sheared to give a tight size distribution (approx 2 kb).  
 The v + i method used for the library construction is  
 described in detail in Smith, H.O. and Venter, J.C.

Seq primer: M13-Forward  
 Class: shotgun.

FEATURES

Source

Location/Qualifiers

1..293

/organism="Trypanosoma brucei"

/strain="TREU927/4 Gutat 10.1"

/db\_xref="taxon:5691"

/clone="Sheared DNA-67G1"

/clone\_lib="Sheared DNA"

/note="Vector: pUC18; Site:1: SmaI; Constructed at The  
 Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from a cloned population of  
 Trypanosoma brucei (TREU927/4 Gutat 10.1) was mechanically  
 sheared to give a tight size distribution (approx 2 kb).  
 The v + i method used for the library construction is  
 described in detail in Smith, H.O. and Venter, J.C.

/db\_xref="taxon:10090"  
 /clone="IMAGE:1038965"  
 /clone\_lib="Stratagene mouse heart (#937316)"  
 /sex="pooled"  
 /tissue\_type="heart"  
 /dev\_stage="13 day embryos"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: heart; Vector: pBluescript SK-; Site:1:  
 EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:  
 Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts.  
 Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'  
 adaptor sequence: 5' GAATTCGCACGAG 3' -3' adaptor  
 sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'

BASE COUNT 26 a 25 c 25 g 30 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 23; Length 106;  
 Best Local Similarity 87.0%; Pred. No. 7.8e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgctggtatctgtgttgaagc 24  
 |||||

DB 84 TTGCTGGATCTGCCGTTGGAGC 106

#### RESULT 18

LOCUS AZ214122 293 bp DNA 09-JUN-2000  
 DEFINITION Sheared DNA-67G1.TF Sheared DNA Trypanosoma brucei genomic clone  
 Sheared DNA-67G1, DNA sequence.

ACCESSION AZ214122  
 VERSION AZ214122.1 GI:8431922  
 KEYWORDS GSS.  
 SOURCE Trypanosoma brucei.  
 ORGANISM Trypanosoma brucei  
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
 Trypanosoma.

REFERENCE 1 (bases 1 to 293)  
 AUTHORS El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,  
 Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,  
 Fraser, C. and Adams, M.

TITLE Determination of clone end sequences from Trypanosoma brucei GUTat  
 10.1 sheared DNA library

JOURNAL Unpublished (1999)  
 COMMENT Other\_GSSs: Sheared DNA-67G1.TR  
 Contact: Najib M. El-Sayed  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208

Email: nelsayed@tigr.org  
 Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared  
 DNA library constructed at TIGR. Clones will be available for  
 distribution through Research Genetics, Alabama, USA. Sheared DNA  
 end sequences search page: <http://www.tigr.org/tdb/mdb/tbdb/>.

Seq primer: M13-Forward  
 Class: shotgun.

FEATURES

Location/Qualifiers

1..293

/organism="Trypanosoma brucei"

/strain="TREU927/4 Gutat 10.1"

/db\_xref="taxon:5691"

/clone="Sheared DNA-67G1"

/clone\_lib="Sheared DNA"

/note="Vector: pUC18; Site:1: SmaI; Constructed at The  
 Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from a cloned population of  
 Trypanosoma brucei (TREU927/4 Gutat 10.1) was mechanically  
 sheared to give a tight size distribution (approx 2 kb).  
 The v + i method used for the library construction is  
 described in detail in Smith, H.O. and Venter, J.C.

Seq primer: M13-Forward  
 Class: shotgun.

FEATURES

Location/Qualifiers

1..293

/organism="Trypanosoma brucei"

/strain="TREU927/4 Gutat 10.1"

/db\_xref="taxon:5691"

/clone="Sheared DNA-67G1"

/clone\_lib="Sheared DNA"

(Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."

BASE COUNT 38 a 73 c 78 g 104 t  
ORIGIN

Query Match 72.8%; Score 18.2; DB 240; Length 293;  
Best Local Similarity 87.0%; Pred. No. 8.9e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgtggtatcgtgtttgaagc 24  
||||| ||||||| ||||||| |||||||  
Db 142 TTGGTTATCTGCTGTTGAAGC 164

RESULT 19  
LOCUS BE990318 303 bp mRNA EST 05-OCT-2000  
DEFINITION UI-M-BZ1-bfu-n-15-0-UI-s1 NIH\_BMAP\_MHI2\_S1 Mus musculus cDNA clone  
UI-M-BZ1-bfu-n-15-0-UI 3', mRNA sequence.

ACCESSION BE990318  
VERSION BE990318.1 GI:10668680  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 303)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 9704477  
COMMENT

Contact: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mest@mail.nih.gov

The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to verify it as a clone from the  
hippocampus tissue cDNA Library preparation: M.B. Soares Lab Clone  
distribution: Researchers may obtain BMAP cDNA clones from RESEARCH  
GENETICS. It should be noted that Bento Soares is generating a  
small number of additional specialized non-redundant arrays of BMAP  
cDNAs whose availability will be considered under appropriate and  
limited collaborative arrangements The following repetitive  
elements were found in this cDNA sequence: 1-22,  
>AT-rich#Low complexity  
Seq primer: M13 Forward

POLYA=Yes.  
Location/Qualifiers  
1..303

FEATURES  
source

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-BZ1-bfu-n-15-0-UI"  
/clone\_lib="NIH\_BMAP\_MHI2\_S1"  
/lab\_host="27-32 days"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The  
NIH\_BMAP\_MHI2\_S1 library is a subtracted library derived  
from NIH\_BMAP\_MHI2. NIH\_BMAP\_MHI2 is a library derived  
from mouse hippocampus tissue. For a detailed description  
of the library from which this clone was derived, please  
visit our web site at brainest.eng.uiowa.edu.

TAG\_LIB=NIH\_BMAP\_MHI2\_S1  
TAG\_TISSUE=hippocampus  
TAG\_SEQ=RAGTC

BASE COUNT 78 a 77 c 60 g 88 t  
ORIGIN

Query Match 72.8%; Score 18.2; DB 143; Length 303;  
Best Local Similarity 87.0%; Pred. No. 8.9e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 tgctggtatcgtgtttgaagc 25  
||||| ||||||| ||||||| |||||||  
Db 263 TGCTGATCTGCTCTCATCG 285

RESULT 20  
LOCUS BG008569 345 bp mRNA EST 24-JAN-2001  
DEFINITION PM4-GN0304-291100-001-c07 GN0304 Homo sapiens cDNA, mRNA sequence.  
BG008569  
ACCESSION BG008569.1 GI:12453896  
VERSION BG008569.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 345)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM4&st2=PM4-GN0304-  
291100-001-c07&t3=2000-11-29&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 14  
High quality sequence stop: 345.  
Location/Qualifiers  
1..345

FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0304"  
/dev\_stage="Adult"

FEATURES  
source

Location/Qualifiers  
1..345

FEATURES  
source

Location/Qualifiers  
1..345

FEATURES  
source

Location/Qualifiers  
1..345

FEATURES  
source

Location/Qualifiers  
1..345

FEATURES  
source

Location/Qualifiers  
1..345

FEATURES  
source

Location/Qualifiers  
1..345

```

QY 2 ttgttgatctgtgtttgaagc 24
||||| ||||||| |||
Db 242 TTGCTGGCTCTGCTGTTTCAGC 220

RESULT 21
AZ412964/c 422 bp DNA GSS 03-OCT-2000
LOCUS 1M0186F13R Mouse 10kb plasmid UUGCLM library Mus musculus genomic
DEFINITION clone UUGCLM0186F13 R, DNA sequence.
ACCESSION AZ412964
VERSION AZ412964.1 GI:10536977
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 422)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0186 row: F column: 13
Seq primer: CACACAGGAAACAGTATGACC
Class: plasmid ends
High quality sequence stop: 422.
FEATURES
source
Location/Qualifiers
1..422
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0186F13"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (g1473214[gblAFI29072.1]), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 104 a 92 c 90 g 136 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 243; Length 422;
Best Local Similarity 87.0%; Pred. No. 9.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgttgatctgtgtttgaagc 24
||||| ||||||| |||
Db 242 TTGCTGGCTCTGCTGTTTCAGC 220

RESULT 22
A1449930/c 437 bp mRNA EST 09-MAR-1999
LOCUS mr81a05.x1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
DEFINITION IMAGE:603824 3', mRNA sequence.
ACCESSION A1449930
VERSION A1449930.1 GI:4293448
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 437)
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person,
B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter,
E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R., and Wilson, R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
High quality sequence stop: 393.
FEATURES
source
Location/Qualifiers
1..437
/organism="Mus musculus"
/strain="NIH/Swiss"
/db_xref="taxon:10090"
/clone="IMAGE:603824"
/clone_lib="Stratagene mouse heart (#937316)"
/sex="pooled"
/tissue_type="heart"
/dev_stage="13 day embryos"
/lab_host="SOLR (kanamycin resistant)"
/notes="Organ: heart; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dr. 93 pooled NIH/Swiss 13 day embryo hearts.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5,
adaptor sequence: 5' GAATTCGCACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT 121 a 76 c 99 g 140 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 20; Length 437;
Best Local Similarity 87.0%; Pred. No. 9.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgttgatctgtgtttgaagc 24
||||| ||||||| |||
Db 31 TTGCTGGATCTGCCGTTGGAGC 9

RESULT 23
A0522448 437 bp DNA GSS 11-MAY-1999
LOCUS HS_5237_A2_A12_77A RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate-813 Col=24 Row=A, DNA sequence.

```





/tissue\_type="colon"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from 12 pooled bulk tumor samples and primed  
with a Not I - oligo(dT) primer. Double-stranded cDNA was  
ligated to Eco RI adaptors (Pharmacia), digested with Not  
I and cloned into the Not I and Eco RI sites of the  
modified pT73 vector. Library went through one round of  
normalization."  
BASE COUNT 177 a 89 c 71 g 158 t  
ORIGIN

Query Match 72.8%; Score 18.2; DB 8; Length 495;  
Best Local Similarity 87.0%; Pred. No. 9.5e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 gttgctggtctgctgtttgaag 23  
||||| ||||| || ||||| |||||  
Db 451 GTTGCTGGACCTACTGTCGAAG 429

RESULT 26  
AW702183  
LOCUS  
DEFINITION AW702183 499 bp mRNA EST 22-MAY-2000  
TgESTz73b05.y1 TgRH\*-Tachyzoite cDNA Toxoplasma gondii cDNA clone  
TgESTz73b05.y1 5', mRNA sequence.  
ACCESSION AW702183  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Toxoplasma gondii.  
Toxoplasma gondii.  
Eukaryota; Alveolata; Apicomplexa; Coccidia; Elmeriida;  
Sarcocystidae; Toxoplasma.  
REFERENCE  
AUTHORS  
Hehl, A., Manger, I., Marra, M., Sibley, L.D., Ajioke, J.A., Aslett, M.A.,  
Dietrich, N., Dubucque, T., Hillier, L., Kucaba, T., Wan, K.L.,  
Waterson, R.H., and Boothroyd, J.  
WashU-Merck-Stanford-NIH Toxoplasma EST project  
Unpublished (1996)  
Contact: Marra M  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: toxowatson.wustl.edu  
Contact David Sibley (toxowest@borcim.wustl.edu) for further  
information relating to organism, libraries, or clone availability.  
Seq primer: -40RP from Gibco  
High quality sequence stop: 415.  
Location/Qualifiers  
1..499  
/organism="Toxoplasma gondii"  
/strain="RH (Type I)"  
/db\_xref="taxon:5811"  
/clone="TgESTz73b05.y1"  
/clone\_lib="TgRH\*-Tachyzoite cDNA"  
/dev\_stage="tachyzoite"  
/lab\_host="SOLR cells"  
/note="Vector: pTachyzoite SK-; Site\_1: EcoRI; Site\_2:  
XhoI; cDNA was synthesized from poly mRNA using an  
oligo-dT primer containing an XhoI site. Following second  
strand synthesis, EcoRI adaptors were ligated to the cDNA  
and products were size-selected on sephacryl S500. The  
cDNAs were ligated to EcoRI/ XhoI prepared lambda ZapII  
(Stratagene). Clones were converted to phagemids by mass  
excision using ExAssist helper phage and E.coli SOLR cells  
(Stratagene). Insert sizes range from 0.3-3.0 kb. The  
library may contain a small percentage of host or  
bacterial contaminants. NOTE: Many clones lack the 5'  
EcoRI site but contain the upstream polylinker sites."

FEATURES  
source  
1..507  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-36K23"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site\_1:  
EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size  
selected DNA was cloned into the pBACE3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."  
BASE COUNT 145 a 108 c 98 g 156 t  
ORIGIN

Query Match 72.8%; Score 18.2; DB 240; Length 507;  
Best Local Similarity 87.0%; Pred. No. 9.5e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 gttgctggtctgctgtttgaag 23

BASE COUNT 129 a 131 c 107 g 130 t 2 others  
ORIGIN  
Query Match 72.8%; Score 18.2; DB 119; Length 499;  
Best Local Similarity 87.0%; Pred. No. 9.5e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 ttgctggtctgctgtttgaag 24  
||||| ||||| || ||||| |||||  
Db 178 TTGCTGGTCTGCTTTTCAAGC 200

RESULT 27  
AZ244099  
LOCUS  
DEFINITION AZ244099 507 bp DNA GSS 15-JUN-2000  
RPCI-23-36K23.TJ RPCI-23 Mus musculus genomic clone RPCI-23-36K23,  
DNA sequence.  
ACCESSION AZ244099  
VERSION AZ244099.1 GI:8557290  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus.  
house mouse.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS  
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret  
B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.  
and Fraser, C.M.  
Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Other\_GSSs: RPCI-23-36K23.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)  
or from Resea ch Genetics (info@resgen.com). BAC end page:  
[http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 36 row: K column: 23  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source  
1..507  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-36K23"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site\_1:  
EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size  
selected DNA was cloned into the pBACE3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."  
BASE COUNT 145 a 108 c 98 g 156 t  
ORIGIN

Query Match 72.8%; Score 18.2; DB 240; Length 507;  
Best Local Similarity 87.0%; Pred. No. 9.5e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 gttgctggtctgctgtttgaag 23

Db 290 GTTCTGCACTGCTGTGTGAAG 312  
|||||

## RESULT 28

BE656252

LOCUS

DEFINITION

BE656252 517 bp mRNA EST 06-SEP-2000  
UI-M-BH0-ajr-b-02-0-UI.r1 NIH\_BMAP\_M.S1 Mus musculus cDNA clone  
UI-M-BH0-ajr-b-02-0-UI 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

1 (bases 1 to 517)

Ronaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

9704477

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mestr@mail.nih.gov

cDNA Library preparation: M.B. Soares Lab Clone distribution:

Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It

should be noted that Bento Soares is generating a small number of

additional specialized non-redundant arrays of BMAP cDNAs whose

availability will be considered under appropriate and limited

collaborative arrangements. The following repetitive elements were

found in this cDNA sequence: 10-93, &gt;(GGAA)n#Simple\_repeat

Seq primer: M13 Reverse.

FEATURES

source

1..517

Location/Qualifiers

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UI-M-BH0-ajr-b-02-0-UI"

/clone\_lib="NIH\_BMAP\_M.S1"

/dev\_stage="27-32 days"

/lab\_host="DH10B (Life Technologies)"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker; Site\_1: Not I; Site\_2: Eco RI; The

NIH\_BMAP\_M.S1 library is a subtracted library derived from

a mixture of normalized libraries from ten regions of the

mouse brain (cerebellum, brain stems, olfactory bulbs,

hypothalamus, cortex, amygdala, basal ganglia, pineal

gland, striatum, hippocampus). The driver used for

subtraction consisted of a pool of 20,000 cDNA clones

obtained from non-normalized and normalized libraries of

these ten regions of the mouse brain."

BASE COUNT 114 a 150 c 106 g 147 t

ORIGIN

Query Match

Best Local Similarity 72.8%; Score 18.2; DB 138; Length 517;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgctggtatctgtgttgaagc 24

|||||

Db 266 TTGCTGGACTGATGTTTGAAGC 288

RESULT 29

BE443324/c

LOCUS

DEFINITION

BE443324 521 bp mRNA EST 25-JUL-2000  
WHE1112\_E04\_J082S Wheat etiolated seedling root normalized cDNA

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae

; Triticeae; Triticum.

1 (bases 1 to 521)

Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han

, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Nguyen, H.T.,

Rausch, C.J., Seaton, C.L., Tong, J.C. and Zhang, D.

The structure and function of the expressed portion of the wheat

genomes - Normalized root cDNA library

Unpublished (2000)

Contact: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific

West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773

Fax: 5105595818

Email: oanderson@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low

quality sequence with phred score less than 20

Seq primer: Stratagene SK primer.

FEATURES

source

1..521

Location/Qualifiers

/organism="Triticum aestivum"

/cultivar="Chinese Spring"

/db\_xref="taxon:4565"

/clone="WHE1112\_E04\_J08"

/clone\_lib="Wheat etiolated seedling root normalized cDNA

library"

/tissue\_type="Root"

/dev\_stage="Five day old etiolated seedling"

/lab\_host="E. coli DH10B"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid

pBluescript SK; Site\_1: EcoRI; Site\_2: XhoI; Seeds were

surface-sterilized, germinated and grown aseptically in

the dark at room temperature on filter paper with water,

nystatin and cefotaxime in covered crystallization

dishes. Roots were harvested. The tissue, total RNA, and

poly(A) RNA were prepared, a cDNA library was made in the

TJ Close lab (Choi, Close, Fenton) at the University of

California, Riverside. The cDNA clones were in vivo

excised to give pBluescript phagemids before

normalization was carried out. The mass excision of

phagemid library and normalization were done in HT Nguyen

lab by D. Zhang at Texas Tech University. Normalization

protocol used was that of Soares. Plasmid DNA

preparations and DNA sequencing were performed in the OD

Anderson lab (all other authors)."

BASE COUNT 127 a 110 c 89 g 195 t

ORIGIN

Query Match

Best Local Similarity 72.8%; Score 18.2; DB 167; Length 521;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 tgcctggtatctgtttgaagc 25

|||||

Db 64 TCCTGCACTGCTGTGTTGAAGC 42

RESULT 30

TA375E11P/c

LOCUS

DEFINITION

TA375E11P 532 bp DNA GSS 13-DEC-2000

T. brucei sheared genomic DNA clone 375e11, forward sequence,

genomic survey sequence.

ACCESSION AL495591

library Triticum aestivum cDNA clone WHE1112\_E04\_J08, mRNA

sequence.

BE443324

VERSION BE443324.1 GI:9442856

KEYWORDS EST.

SOURCE bread wheat.

ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae

; Triticeae; Triticum.

1 (bases 1 to 521)

Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han

, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Nguyen, H.T.,

Rausch, C.J., Seaton, C.L., Tong, J.C. and Zhang, D.

The structure and function of the expressed portion of the wheat

genomes - Normalized root cDNA library

Unpublished (2000)

Contact: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific

West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773

Fax: 5105595818

Email: oanderson@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low

quality sequence with phred score less than 20

Seq primer: Stratagene SK primer.

```

library availability, please contact Pieter de Jong
(pieter@jones.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html
Plate: 372 row: G column: 22
Seq primer: SP6
Class: BAC ends.

FEATURES             location/Qualifiers
     source            1..541
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-372G22"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and partially digested
                     with a combination of EcoRI and EcoRI Methylase. Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI sites. The ligation products were transformed into
                     DH10B electrocompetent cells (BRL Life Technologies).
BASE COUNT      129 a 124 c 154 g 134 t
ORIGIN

Query Match      72.8%; Score 18.2; DB 236; Length 541;
Best Local Similarity 87.0%; Pred. No. 9.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0

QY 1 gttctggatcgtctgtttgaag 23
    ||| ||||| || ||||| |||
Db 184 GTTCTGGATGTCATGTTGAAG 206

RESULT 32
TA162C10P/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

TA162C10P 559 bp DNA GSS 13-DEC-2000
T. brucei sheared genomic DNA clone 162c10, forward sequence,
genomic survey sequence.
AL472440
AL472440.1 GI:11837892
GSS.
Trypanosoma brucei.
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 559)
Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
Melville,S.E., Rajadream,M.A. and Barrell,B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T\_brucei/.

```



JOURNAL  
COMMENT

Unpublished (1999)  
Contact: David Frisch  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 4366  
Fax: 864 656 4293  
Email: dfrisch@CLEMSON.EDU  
3 prime sequence.

FEATURES  
Source

Location/Qualifiers  
1. .628

/organism="Lycopersicon pennellii"  
/db\_xref="taxon:28526"  
/clone="clPTIK17"  
/clone.lib="L. pennellii trichome, Cornell University"  
/tissue\_type="trichome"  
/dev\_stage="mixed stages"  
/lab\_host="SOLR"  
/note="vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; Leaves of various stages were shaken in liquid  
nitrogen, shearing off trichomes. This procedure yielded a  
mixture of cells highly enriched for trichomes, with minor  
contamination by other types of leaf cells."

BASE COUNT  
ORIGIN

184 a 140 c 122 g 182 t

Query Match  
Best Local Similarity

72.8%; Score 18.2; DB 112; Length 628;  
87.0%; Pred. No. 9.8e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgctggatctgctgtttgaagc 24

||||| ||| ||| ||| |||

Db 71 TTGCTGGATGTCGTGTTTAAAGC 49

RESULT 36  
TA69C05P/CLOCUS  
DEFINITION

TA69C05P 650 bp DNA GSS 13-DEC-2000  
T. brucei sheared genomic DNA clone 69c05, forward sequence,  
genomic survey sequence.

ACCESSION AL457556

VERSION AL457556.1 GI:11858782

KEYWORDS GSS.

SOURCE Trypanosoma brucei.

ORGANISM Trypanosoma brucei

Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

REFERENCE 1 (bases 1 to 650)

AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,  
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,  
Melville, S.E., Rajandream, M.A. and Barrell, B.G.

TITLE Direct Submission

## JOURNAL

Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing  
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,  
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and  
nhs@sanger.ac.uk

## COMMENT

Constructed at the Institute for Genomic Research (TIGR),  
Rockville, MD. Genomic DNA isolated from a cloned population of  
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared  
to give a tight size distribution (4 kb). The v + i method used for the library construction is  
described in detail in Smith, H. and Venter, J.C. (Making small  
insert libraries for whole genome shotgun sequencing projects. In  
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org

Details of T. brucei sequencing at the Sanger Centre are available  
at [http://www.sanger.ac.uk/Projects/T\\_brucei/](http://www.sanger.ac.uk/Projects/T_brucei/).

FEATURES  
Source

Location/Qualifiers  
1. .650

/organism="Trypanosoma brucei"

/strain="TREU927"

/db\_xref="taxon:5691"  
/clone="69c05"

BASE COUNT 208 a 186 c 173 g 83 t

## ORIGIN

Query Match 72.8%; Score 18.2; DB 258; Length 650;

Best Local Similarity 87.0%; Pred. No. 9.8e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ttgctggatctgctgtttgaagc 24

||||| ||| ||| ||| |||

Db 364 TTGCTTTATCTGCTGTTGAAGC 342

## RESULT 37

## AI325366/C

## LOCUS

## DEFINITION

AI325366 685 bp mRNA EST 23-DEC-1998  
mi29f01.y1 Soares mouse embryo NDME13.5 14.5 Mus musculus cDNA  
clone IMAGE:464953 5' similar to TR:Q60947 Q60947 MAX-INTERACTING  
TRANSCRIPTIONAL REPRESSOR. ;, mRNA sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 685)

AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

TITLE The WashU-HMI Mouse EST Project

UNPUBLISHED (1996)

CONTACT: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

MGI:278769

This read is a RESEQUENCE of a previously sequenced mouse clone

This read has been verified (found to hit its original self in the

correct orientation)

Possible reversed clone: similarity on wrong strand

Seq primer: -40RP from Gibco

High quality sequence stop: 384.

Location/Qualifiers

1. .685

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:464953"

/clone.lib="Soares mouse embryo NDME13.5 14.5"

/sex="unknown"

/tissue\_type="embryo"

/dev\_stage="13.5-14.5dpc total fetus"

/lab\_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTTTTTTTT

T 3'], on equal amounts of mRNA from 2 13.5dpc and 2

14.5dpc embryos [total RNA provided by Minoru Ko, Wayne

State Univ., from 2 ]; double-stranded cDNA was ligated to

Eco RI adaptors (Pharmacia), digested with Not I and

cloned into the Not I and Eco RI sites of the modified

pT7T3 vector. Library went through one round of







## RESULT 42

FR0044714  
 LOCUS  
 DEFINITION Fugu rubripes GSS sequence, clone 192G14eB6, genomic survey sequence.  
 ALI32206  
 VERSION ALI32206.1 GI:6114152  
 KEYWORDS GSS; genome survey sequence.  
 SOURCE Takifugu rubripes  
 ORGANISM Takifugu rubripes  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu.  
 1 (bases 1 to 177)  
 Elgar, G., Clark, M.S., Smith, S., Meek, S., Warner, S., Edwards, Y.J.K., Umranta, Y., Williams, G. and Brenner, S.  
 Direct Submission  
 Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB. UK Email: biohelp@hmp.mrc.ac.uk  
 Vector: pBluescript II KS  
 V type: phagemid  
 PRIMER: KS

DESCR:  
 One pass dye-terminator sequencing of cosmid cloned genomic sequence.

## FEATURES

Source  
 1..177  
 /organism="Takifugu rubripes"  
 /db\_xref="taxon:31033"  
 /clone\_lib="cosmid 192G14"  
 /clone="192G14eE6"

BASE COUNT 33 a 41 c 50 g 44 t 9 others  
 ORIGIN

Query Match 71.2%; Score 17.8; DB 222; Length 177;

Best Local Similarity 90.5%; Pred. No. 1.2e+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 tctgtggtctgtctgttgaag 23  
 | ||| ||||| ||||| |||||  
 Db 110 TCTGAATCTGCTGTGAAG 130

## RESULT 43

AU074974/C  
 LOCUS  
 DEFINITION AU074974 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium discoideum cDNA clone SSM625, mRNA sequence.  
 ACCESSION AU074974  
 VERSION AU074974  
 KEYWORDS EST.  
 SOURCE Dictyostelium discoideum.  
 ORGANISM Dictyostelium discoideum  
 Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.  
 1 (bases 1 to 240)  
 Urushihara, H.  
 Developmental cDNA in Dictyostelium discoideum (1999)  
 Unpublished (1999)  
 Contact: Hideko Urushihara  
 Institute of Biological Sciences  
 University of Tsukuba  
 3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan  
 Email: d402huesakura.cc.tsukuba.ac.jp  
 PROJECT = Dictyostelium discoideum cDNA project in Japan.  
 1..240  
 Location/Qualifiers  
 /organism="Dictyostelium discoideum"  
 /strain="AX4"

## FEATURES

Source  
 1..240  
 /organism="Dictyostelium discoideum"  
 /strain="AX4"

/db\_xref="taxon:44689"  
 /clone="SSM625"  
 /clone\_lib="Dictyostelium discoideum SS (H.Urushihara)"  
 /dev\_stage="slug"

BASE COUNT 88 a 49 c 26 g 76 t 1 others  
 ORIGIN

Query Match 71.2%; Score 17.8; DB 107; Length 240;

Best Local Similarity 90.5%; Pred. No. 1.3e+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gttgctggtctgtctgttga 21  
 | ||||| ||||| ||||| |||||  
 Db 206 GTTGTCTGATTGCTGTTGA 186

## RESULT 44

BF440174  
 LOCUS  
 DEFINITION BF440174 295 bp mRNA EST 01-DEC-2000  
 BS29000185303 Lewin Cattle Spleen Bos taurus cDNA clone  
 BS29000185303 3', mRNA sequence.  
 ACCESSION BF440174  
 VERSION BF440174.1 GI:11500057  
 KEYWORDS EST.  
 SOURCE COW.

ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovidae; Bovinae; Bos.  
 1 (bases 1 to 295)

REFERENCE  
 AUTHORS Band, M.R., Larson, J.H., Reibel, M., Green, C.A., Heyen, D.W., Donovan, J., Windish, R., Steining, C., Mahyuddin, P., Womack, J. and Lewin, H.A.

An ordered comparative map of the cattle and human genomes

Genome Res. 10 (9), 1359-1368 (2000)

2042530

## COMMENT

Contact: Lewin, H. A.

W. M. Keck Center for Comparative and Functional Genomics

University of Illinois at Urbana-Champaign

340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL

61801, USA

Tel: 217 333 5998

Fax: 217 244 5617

Email: h-lewin@uiuc.edu

Funding for Cattle EST sequencing was provided by the USDA National Research Initiative, Project No. 98-35205-6644, and a grant from the Japanese Ministry of Agriculture Fisheries and Forestry to H.A.Lewin and J.E.Womack. Base-calling/Quality scores: PHRED or

Washington University Genome Center.Vector-trimming: Cross-Match from Washington University Genome Center PHRAP suite. This sequence

is vector free and at least 150bp in length.

Insert Length: 295 Std Error: 0.00

Seq primer: TACGACTCACTATAGCGCAAT

High quality sequence stop: 295.

Location/Qualifiers

## FEATURES

Source

1..295

/organism="Bos taurus"

/strain="Angus"

/db\_xref="taxon:9913"

/clone="BS29000185303"

/clone\_lib="Lewin Cattle Spleen"

/sex="Female"

/dev\_stage="Adult"

/note="Organ: Spleen; Vector: pBluescript SK(+/-)"

77 a 70 c 75 t 3 others

## ORIGIN

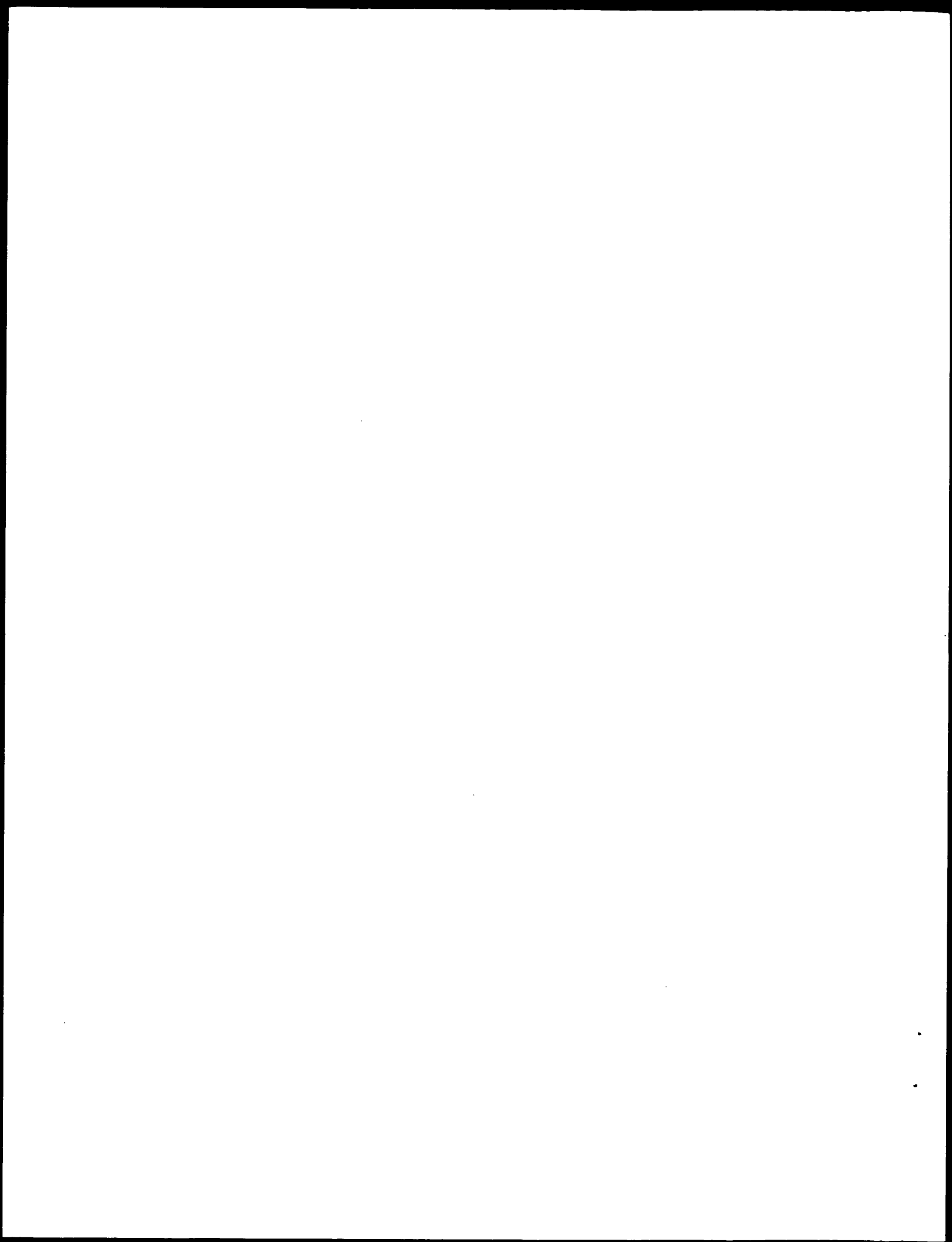
Query Match

Best Local Similarity 71.2%; Score 17.8; DB 148; Length 295;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 12:10:31 ; Search time 1666.31 Seconds  
(without alignments)  
232.066 Million cell updates/sec

Title: US-09-396-196F-4  
Perfect score: 25  
Sequence: 1 ggatctgctgttgaagcgcagcag 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_bal.\*  
2: gb\_ba2.\*  
3: gb\_ba3.\*  
4: gb\_in1.\*  
5: gb\_in2.\*  
6: gb\_in3.\*  
7: gb\_om.\*  
8: gb\_ov.\*  
9: gb\_pat1.\*  
10: gb\_pat2.\*  
11: gb\_ph.\*  
12: gb\_pl1.\*  
13: gb\_pl2.\*  
14: gb\_pl3.\*  
15: gb\_pl4.\*  
16: em\_bal.\*  
17: em\_ba2.\*  
18: em\_fun.\*  
19: em\_higo\_hum.\*  
20: em\_higo\_inv.\*  
21: em\_higo\_rod.\*  
22: em\_hig\_hum1.\*  
23: em\_hig\_hum2.\*  
24: em\_hig\_hum3.\*  
25: em\_hig\_hum4.\*  
26: em\_hig\_hum5.\*  
27: em\_hig\_hum6.\*  
28: em\_hig\_hum7.\*  
29: em\_hig\_hum8.\*  
30: em\_hig\_inv1.\*  
31: em\_hig\_inv2.\*  
32: em\_hig\_other.\*  
33: em\_hig\_rod.\*  
34: em\_hum1.\*  
35: em\_hum2.\*  
36: em\_hum3.\*  
37: em\_hum4.\*  
38: em\_hum5.\*  
39: em\_hum6.\*  
40: em\_hum7.\*  
41: em\_in.\*  
42: em\_om.\*  
43: em\_or.\*

44: em\_ov.\*  
45: em\_pat.\*  
46: em\_ph.\*  
47: em\_pl.\*  
48: em\_ro.\*  
49: em\_sl.\*  
50: em\_sy.\*  
51: em\_un.\*  
52: em\_v1.\*  
53: gb\_sts1.\*  
54: gb\_sts2.\*  
55: gb\_sts3.\*  
56: gb\_sy.\*  
57: gb\_un.\*  
58: gb\_v11.\*  
59: gb\_v12.\*  
60: gb\_htg1.\*  
61: gb\_htg2.\*  
62: gb\_htg3.\*  
63: gb\_htg4.\*  
64: gb\_htg5.\*  
65: gb\_htg6.\*  
66: gb\_htg7.\*  
67: gb\_htg8.\*  
68: gb\_htg9.\*  
69: gb\_htg10.\*  
70: gb\_htg11.\*  
71: gb\_htg12.\*  
72: gb\_htg13.\*  
73: gb\_htg14.\*  
74: gb\_htg15.\*  
75: gb\_htg16.\*  
76: gb\_htg17.\*  
77: gb\_htg18.\*  
78: gb\_htg19.\*  
79: gb\_htg20.\*  
80: gb\_htg21.\*  
81: gb\_htg22.\*  
82: gb\_htg23.\*  
83: gb\_htg24.\*  
84: gb\_htg25.\*  
85: gb\_pr1.\*  
86: gb\_pr2.\*  
87: gb\_pr3.\*  
88: gb\_pr4.\*  
89: gb\_pr5.\*  
90: gb\_pr6.\*  
91: gb\_pr7.\*  
92: gb\_pr8.\*  
93: gb\_pr9.\*  
94: gb\_ro1.\*  
95: gb\_ro2.\*  
96: gb\_in4.\*  
97: gb\_pr10.\*  
98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	25	100.0	1041	9 AR029499	Sequence
2	25	100.0	1041	9 AR034916	Sequence
3	25	100.0	1084	9 A11530	BioB gene o
4	25	100.0	5526	2 AF250776	Unculture
5	25	100.0	5793	2 ECOBIO	E.coli 7, 8-
6	25	100.0	5872	9 A38246	Sequence 1
7	25	100.0	5872	9 A38251	Sequence 6
8	25	100.0	5872	9 A93674	Sequence 1

```

9 25 100.0 5872 9 A93679 Sequence 6
10 25 100.0 5872 9 AR101809 Sequence
11 25 100.0 5872 9 AR101810 Sequence
12 25 100.0 11022 1 AE000180 Escherich
13 25 100.0 12891 1 AE0004192 Vibrio ch
14 25 100.0 13501 1 AE005258 Escherich
15 25 100.0 297816 2 AP002553 Escherich
16 23.4 93.6 1121 10 E00893 Escherich
17 22.4 89.6 8227 2 AF248314 Genomic DNA
18 21 84.0 128 2 E00893 Genomic DNA
19 19.8 79.2 80993 4 AC004313 Unculture
20 19.8 79.2 83551 65 AC004313 Escherichia
21 19.4 77.6 1041 3 AC018150 Drosophil
22 19.2 76.8 86446 63 AC018150 Drosophil
23 19.2 76.8 182525 4 AC009213 Erwinia her
24 19.2 76.8 197348 4 AC009213 Drosophil
25 19.2 76.8 239171 5 AC008028 Drosophil
26 18.8 75.2 2472 94 AF052042 Drosophil
27 18.8 75.2 8371 1 AF052042 Rattus no
28 18.8 75.2 16147 6 AF087452 Bacillus
29 18.8 75.2 190673 4 AC008357 Drosophil
30 18.8 75.2 191736 69 AC008357 Drosophil
31 18.8 75.2 221888 5 AC025411 Homo sapi
32 18.6 74.4 1610 58 AB014726 Drosophil
33 18.6 74.4 2191 89 AB014726 Chuzan vi
34 18.6 74.4 4501 2 AF188739 Homo sapi
35 18.6 74.4 6572 94 AF188739 Homo sapi
36 18.6 74.4 10529 3 M69036 Alcaligenes
37 18.6 74.4 11208 3 X57144 H. halobacte
38 18.6 74.4 39373 84 AE005138 Halobacte
39 18.6 74.4 70326 73 AC068957 Homo sapien
40 18.6 74.4 150516 67 AC022946 Homo sapi
41 18.6 74.4 151696 89 AF240629 Homo sapi
42 18.6 74.4 154923 78 AF240629 Homo sapi
43 18.6 74.4 170896 61 AC011010 Homo sapi
44 18.6 74.4 190937 91 AP000957 Homo sapi
45 18.6 74.4 201652 64 AC016821 Homo sapi

ALIGNMENTS

RESULT 1
AR029499
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
AUTHORS Patton,D.Andrew
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES Location/Qualifiers
source
1..1041
/organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatctgctgtttgaagcgcagcag 25
Db 63 GGATCTGCTGTTGAAGCGCAGCAG 87

RESULT 2
AR034916
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES Location/Qualifiers
source
1..1041
/organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatctgctgtttgaagcgcagcag 25
Db 63 GGATCTGCTGTTGAAGCGCAGCAG 87

RESULT 3
AR1530
LOCUS AR1530 1084 bp DNA
DEFINITION BioB gene of E.coli with primers.
ACCESSION AR1530
VERSION AR1530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli
REFERENCE 1 (bases 1 to 1084)
AUTHORS
JOURNAL Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES Location/Qualifiers
source
1..1084
/organism="Escherichia coli"
/db_xref="taxon:562"
24..1064
/gene="BioB"
24..1064
/gene="BioB"
/codon_start=1
/transl_table=11
/protein_id="CAA00965.1"
/db_xref="GI:490219"
/db_xref="SWISS-PROT:P12996"
/translacion="NAHPRWTLSTQVTELEFKPLDLLFEAQOVHROHFDPRQVQVST
LLSIKIGACPEDCKYCPQSSRYKLGLEAEERLMEVEQVLESARKAKAAGSTRFCMAAW
KNPHERDMPYLEQWQGVKAMGLEACMTLGTLSQOARLANAGLDYNNHLDTSPEF
YGNITITTYQERLDLTLEKVGSDAGIKVCGSIGVIGETVKDRGLLQLLANLPTPES
VPINMLVKVGTPIADNDVDADFIRPT IAVARIIMPTSYVRLSAGREOMNEOTQAMC
FMAGANSIFYGCKLITTPENPEDKDLQFLKRLGLNPQQTAVLAGDNEQQQRLEQALMT
PDTDEYNAAL"
BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatctgctgtttgaagcgcagcag 25
Db 63 GGATCTGCTGTTGAAGCGCAGCAG 87

```

Db 86 GGATCTGCTGTTTGAACGCCGACGAG 110

RESULT 4  
AF250776  
LOCUS  
DEFINITION  
AF250776 5526 bp DNA BCT 31-JAN-2001  
Uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in  
modC-bioA intergenic region, DAPA-aminotransferase BioA (bioA),  
biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and  
biotin biosynthesis protein BioC (bioC) genes, complete cds; and  
dethiobiotin synthetase BioD (bioD) gene, partial cds.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

1. 5526  
/organism="uncultured bacterium pCosHE2"  
/db\_xref="taxon:143797"  
/clone="pCosHE2"  
/note="unknown organism, cosmid clone derived from  
environmental consortium"  
/note="ORF1"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical 17.1 kDa protein in modC-bioA  
intergenic region"  
/protein\_id="AAG60577.1"  
/db\_xref="GI:12620125"  
/translation="MKLSINDLRDGLPHRRHFNMGVYDGNISPLHAWDDVPACTK  
SVVTCYDPDPTGSGMHWVNLFPADTRVLPQGFSGVLAMPDGVLTQRTDFGRAG  
YDGAAPPKGTHTYIFTHALDIERIDVDDEGAGMVGYNVHFHSLASITAMFS"  
/gene="bioA"  
/complement(587..1876)  
/gene="bioA"  
/complement(587..1876)  
/note="bioA"  
/note="7,8-diaminopelargonic acid  
synthetase-aminotransferase"  
/codon\_start=1  
/transl\_table=11  
/product="DAPA-aminotransferase BioA"  
/protein\_id="AAG60578.1"  
/db\_xref="GI:12620126"  
/translation="MTTDDLAFDQRHWHIHPYTSMTSLPYPVYVVSAGCEILSDGRR  
LVDCSSWMAIHGYNHNPALNAAMKSOIDAMSHVMFGGITHPAIELCRKLVAIMPDP  
LKCVELADSGSVAEVMKALQYQAKGAEORFLTFRNGHDTFGAMSVCDPDNS  
MHSWKGYLPENLPAPOSMGDEWDERDMVGFARLMAARHEIAAVILIEVQVAG  
GMWYHPWLKIRKICDREGLIIADIEATGFRNGKLFACAEHAETAPDILGLKAL  
TGGPTMTLSATLTTRAEITISNGEACFMHGPFTFMGPLACAAANASATILESQWQ  
QVADIEVQLREQAPARDAEMVADYVIGVETHTHPYVMAALQKFFVEQGVWIRP  
FGKLIYLMPPYIILLPOOLRLTAAVNRVAVQDETFQC"

1963..3003  
/gene="bioB"  
1963..3003  
/gene="bioB"  
/codon\_start=1  
/transl\_table=11  
/product="biotin synthase BioB"

protein\_id="AAG60579.1"  
/db\_xref="GI:12620127"  
/translation="MAHRPRTLSQVTELFKEKPLLDLLFRAQQVHROHEDPRQVQVST  
LLSXTGACPEDCKYCPOSSRYKTGLEAERLMEVEQVLESARKAKAGSTFRFCMGAAW  
KNPRMDPYLEQMDVQVKGDLGLACMTGLTGLSEQAORLAKGLLQNLNLPPTPEF  
YGNITITTRQERLDTLEKVRDAGIKVCSGIGVGETVKDRAGLLQLLANLPPTPEF  
PINLVAVKGTPIADNDVDVADFIRTIIVARIMPTSYVRLSAGREQNEQOAMC  
FMAGANSIFYGCKLLITPNPEEDKDLQFLKGLGNPQQTAVLAGDNEQQORLEQALMT  
PDTEYNNAAAL"  
3000..4154  
/gene="bioF"  
3000..4154  
/gene="bioF"  
/note="8-amino-7-oxononanoate synthase"  
/codon\_start=1  
/transl\_table=11  
/product="KAPA synthetase BioF"  
/protein\_id="AAG60580.1"  
/db\_xref="GI:12620128"  
/translation="MSWQEKINAAALDARRAADALRRYPVAGAGRWLVADROYLNF  
SSNDYLGSLSHHPQIIRAMQGAEOFGIGSGSGHVSQVSVVHQAEELEAEWIGYSRA  
LLFTSGPAAQVIAAMAKEDRTAADRLSHASLLEAASLSFSQLRFAHNDVTHLAR  
LLASPCQOQIVVTGEGVFSMDGDSAPLAEIQOQVQOHHGWLMDVDAHGTGVIQGGRG  
SCWLQKVPPELLVTFGFGVSGAAVCSSTVADYLLQFAHHLITYSMPAPQAQL  
RASLAVIRDEGDARREKLIVSLIAHFRAGVQDLPETLADSCSAIQPLTVGDSRALQL  
AEKLRQCGCVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"  
4141..4896  
/gene="bioC"  
4141..4896  
/gene="bioC"  
/note="reaction step prior to synthesis of pimeloyl-CoA"  
/codon\_start=1  
/transl\_table=11  
/product="biotin biosynthesis protein BioC"  
/protein\_id="AAG60581.1"  
/db\_xref="GI:12620129"  
/translation="NATVKNQAIATAAFRAALYEQHADLQDSADVLAMLPKRYT  
HVLDAQSGPQWMTWHRERHAQVATLDLSPMLQARQKDAADHYLAGDIESLPLATA  
TFDLANSILAVOMQCNLSALRELYRVRSKVAVFTTVOGSLPELHOAOVDERP  
HANRFLPDEIEQSLNGVYQHIIQPIILWEDDALSAMRSLKIGATHLHEGROPRIIL  
TRSQLRLQALAMPQQQGRYPLTYHLFLGVIARE"  
4883..5526  
/gene="bioD"  
4883..5526  
/gene="bioD"  
/note="DTB synthetase"  
/codon\_start=1  
/transl\_table=11  
/product="dethiobiotin synthetase BioD"  
/protein\_id="AAG60582.1"  
/db\_xref="GI:12620130"  
/translation="MLVSKRYFTGTDTVEGKTVASCALLQAAKAGVYTAGYKPVAS  
GSEKTEPGLRNSDALALQKNSLIQLDYATVNTYTAETPSPHILSAQGRPLESSVMS  
SGLRALQEQADWLVLEGWFTPLSDTFTFADWVYEQOLPVILVVGKLGGINHAMI  
TAQAIOHAGLTLAGWVANDVTPPKRHAHYITTLTRMIPAPLLGEIPMLAENPE"

BASE COUNT 1274 a 1507 c 1567 g 1178 t  
ORIGIN

Query Match 100.0%; Score 25; DB 2; Length 5526;  
Best Local Similarity 100.0%; Pred. No. 0.27;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatgctgtttgaacgccgacgag 25  
|||||

Db 2025 GGATCTGCTGTTTGAACGCCGACGAG 2049

RESULT 5  
ECOBIO  
LOCUS  
DEFINITION  
E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase  
(bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioc









```

/codon_start=1
/transl_table=11
/evidence-experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDRGDKLPHRVFNGMGYDGDNISPHLAWDDVPAGTK
SFVTCVDPDAPTCGWWHVVNLPADTRVLPQFGSLVAMPDGLQTRDFGKTG
YDGAAPKGEHRHIFVTHALDIERIDVDEGASAMVGFNVHSLASAITAMFS"
583..5605
583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgagcagcag 25
|||||
Db 179 GGATCTGCTGTTGAAGCGCAGCAG 203

RESULT 9
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS
DEFINITION Sequence 6 from Patent EP0798384.
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O. and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
Location/Qualifiers
source
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence-experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQKFINAALDARRAADLRRRYPAQAGRWLVADDDROYLNF
SNDVLGLSHHPQILIRAWQGAQEFQIGSGGCHVSGYVVHQALEELAEWLGV SRA
LFTSGFRANQAVIAMMAKEDRIADRLSHASLLEAASLSPSQLRRAHNDVTHLAR
LIASPCPGQWVTEGVFSMDGDSAPLAEIQOVTQOHNGWLMVDDAHGTGVIGEGRG
SCWLQKVLLELVITFGKFGVSGAAVLCSSTVADYLIQFAHLLIYSTMPPAQAL
RASLAVIRSEDEGDAREKLAALITFRAGVQDLPTFLADSCSAIQPLIVGNSRALQL
AEKLQOQGWTAIRPPTVTPAGTARLRLTLTAHEMQDIDRLLEVLHGNG"
3030..3045
/standard_name="BIOF RBS"
3043..3753
/gene="BIOF"
RBS
gene
CDS

```

```

3043..3753
/gene="BIOF"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence-experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRFPVTCGTDEKGTAVASCALLOAKAAGYRTAGYKPVASGS
EKTEPEGRNSDALAQNRSSQLQDYATVMPYFAEPTSPHLSAQGRFIESLWVSAG
LRALEQOADMVLVEGAGWFTPLSDFTTFADWVTOQLPVLVGVKLGCIINHAMLTA
QVIOHAGLTLAGWVANDVPPGKRHAETMTTLTRMIPAPLLGEPWLAENPENAAATGK
YINLAFVDASTIGTSTRL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgagcagcag 25
|||||
Db 179 GGATCTGCTGTTGAAGCGCAGCAG 203

RESULT 10
AR101809 5872 bp DNA PAT 14-FEB-2001
LOCUS
DEFINITION Sequence 1 from patent US 6083712.
ACCESSION AR101809
VERSION AR101809.1 GI:12812607
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
FEATURES
Location/Qualifiers
source
1..5872
/organism="unknown"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgagcagcag 25
|||||
Db 179 GGATCTGCTGTTGAAGCGCAGCAG 203

RESULT 11
AR101810 5872 bp DNA PAT 14-FEB-2001
LOCUS
DEFINITION Sequence 6 from patent US 6083712.
ACCESSION AR101810
VERSION AR101810.1 GI:12812608
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
FEATURES
Location/Qualifiers

```

```

source          1. .5872
                /organism="unknown"
BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match     100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcgtcgtttgaagcagcagcag 25
    |||||
Db 179 GGATCGTCTTTGAAGCGCAGCAG 203

RESULT 12
AE000180
LOCUS          AE000180 11022 bp DNA BCT 01-DEC-2000
DEFINITION     Escherichia coli K12 MGL655 section 70 of 400 of the complete
                genome.
ACCESSION      AE000180 U00096
VERSION        AE000180.1 GI:1786988
SOURCE         Escherichia coli K12.
ORGANISM       Escherichia coli K12.
                Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
                Escherichia.
REFERENCE      1 (bases 1 to 11022)
AUTHORS        Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
                Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
                Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
                Mau, B. and Shao, Y.
                The complete genome sequence of Escherichia coli K-12
                Science 277 (5331), 1453-1474 (1997)
                97426617
PUBLISHED     9778503
REFERENCE      2 (bases 1 to 11022)
AUTHORS        Blattner, F.R.
TITLE         Direct Submission
JOURNAL       Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
                University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
                Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
                608-263-7459
REFERENCE      3 (bases 1 to 11022)
AUTHORS        Blattner, F.R.
TITLE         Direct Submission
JOURNAL       Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
                University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
                Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
                608-263-7459
REFERENCE      4 (bases 1 to 11022)
AUTHORS        Plunkett, G. III.
TITLE         Direct Submission
JOURNAL       Submitted (13-OCT-1998) Laboratory of Genetics, University of
                Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
                This sequence was determined by the E. coli Genome Project at the
                University of Wisconsin-Madison (Frederick R. Blattner, director).
                Supported by NIH grants HG00301 and HG01428 (from the Human Genome
                Project and MCHGR). The entire sequence was independently
                determined from E. coli K12 strain MGL655. Predicted open reading
                frames were determined using GeneMark software, kindly supplied by
                Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
                30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
                have been correlated with genetic loci are being annotated with CG
                Site Nos., unique ID nos. for the genes in the E. coli Genetic
                Stock Center (CGSC) database at Yale University, kindly supplied by
                Mary Berlyn. A public version of the database is accessible
                (http://cgsc.biology.yale.edu). Annotation of the genome is an
                ongoing task whose goal is to make the genome sequence more useful
                by correlating it with other data. Comments to the authors are
                appreciated. Updated information will be available at the E. coli
                Genome Project's World Wide Web site
                (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
FEATURES
source
repeat_region
gene
CDS
protein_bind
gene
CDS
protein_bind
gene
CDS

```

its annotations are periodically updated: this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

Location/Qualifiers

1. .11022

/organism="Escherichia coli K12"

/strain="K12"

/sub\_strain="MGL655"

/db\_xref="taxon:83333"

<1..130

/note="REP (repetitive extragenic palindromic) element; contains 4 REP sequences"

/complement(147..1430)

/gene="ybhC"

/note="b0772"

/complement(147..1430)

/gene="ybhC"

/function="orf; Not classified"

/note="f427; 98 pct identical to fragment YBHC\_ECOLI SW:P46130 (300 aa) but contains 127 additional C-terminal residues"

/codon\_start=1

/transl\_table=11

/product="putative pectinesterase"

/protein\_id="AAC73859.1"

/db\_xref="GI:1786989"

/translation="MNTFSVRLALALAFVTLTACSTPPDQPSDOTAPGTSRRI LSAEAOEDAOHYFASLTPGAANPSPITLPAOPDFVVGPGTQGVHTHTIOAAVD AALIKTKRXYIAVMPGYOCTVYVPAAPGGITLYGTEKPIDVKIGLSDGMSPA DWRHDVNGKYMPCPKAWYDSCSKRSDSIGVCSAFVSWGNLQQLTLT LGSDVAGHPAVALRTDQVQINNINILGRONTFFVNSGVNRLTNRPQRTLVT NSYIEGDVDIVSGRAVDNTEFRVNSRTQOEAYVFAPATLSNIYGFGLAVNRFN AFGDGAQLGRSLDVAANTGQVIRDSAINEGENTAKPWADAVISNRPFGAGTGSVD DNDETQRNLNDTNNRMWEYNRRGVGSKVVAEAKK"

/note="factor Sigma70; predicted +1 start at 806574"

/gene="ybhB"

/note="b0773"

/complement(1582..2058)

/gene="ybhB"

/function="orf; Unknown"

/note="f158; 99 pct identical to YBHB\_ECOLI SW: P12994"

/codon\_start=1

/transl\_table=11

/product="orf, hypothetical protein"

/protein\_id="AAC73860.1"

/db\_xref="GI:1786990"

/translation="MKLISNDLRDGLKPHRFVNGMGYDGNISPHLAWDDVPAGTK SFVWYCYDPDARTGSGWHWVYVNLPAATRVLPQFGSLVAMPDGLQTRDFKGTG YDGAAPPKGETHRIFTVHALDIERIDVDEGASGAMVGFNVHFLSASITAMFS"

/complement(2108..2124)

/note="central position to predicted site"

/bound\_moiety="RhaS predicted site"

/complement(2117..3406)

/gene="bioA"

/note="b0774"

/complement(2117..3406)

/gene="bioA"

/EC\_number="2.6.1.62"

/function="enzyme; Biosynthesis of cofactors, carriers; Biotin"

/note="f429; 100 pct identical to BIOA\_ECOLI SW: P12995"

/codon\_start=1

/transl\_table=11

/product="7,8-diaminopelargonic acid synthetase"

/protein\_id="AAC73861.1"

```
/db_xref="GI:1786991"
/translation="MTTDDLAQDQRHWHPTSWISLPVPVVSAGGCLILISDGR
LECVSSWAAIHYGNPOLNAAMKSOIDAMSHVMEFGIITHAPAIELCRKLVAMTPQ
LVCFLADSGVAEVAMKALQYWOAKGEARQRELTFRNGYHGDTFGAMSVCDPDS
MHSIKCYLLENIPAPAPQSGWDEMDRDMVGFARLMAHREHEIAAVIEPIVQAG
GMRYHPWLKRIKICDREGILLIADEIATGRTGKLFACERAEIADPILCLGKAL
TGGMTLSATLTITREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAIESGDWQ
QVADIEQLREQLAPARDEADVRLVIGAVVETHPVYNNAAALQKFFVEQGVWIRP
FGKLIYLPVPYIILPQQLRLTAANVRVQDETFFCQ"
complement(2193..2221)
/gene="bioA"
/note="factor Sigma70; predicted +1 start at 807260"
3411..3450
/note="central position to bioB promoter: -20"
/bound_moiety="BioB documented site"
3411..3450
/note="central position to predicted promoter: -20; genetic
evidence for the site"
/bound_moiety="BioB predicted site"
complement(3411..3450)
/note="central position to predicted promoter: 50"
/bound_moiety="BioB predicted site"
complement(3411..3450)
/note="central position to bioA promoter: 50"
/bound_moiety="BioB documented site"
3413..3441
/note="factor Sigma70; promoter bioB; documented +1 at
808525"
complement(3447..3473)
/note="factor Sigma70; promoter bioA; documented +1 at
808515"
3493..4533
/gene="bioB"
/note="b0775"
3493..4533
/gene="bioB"
/EC_number="2.8.1.-"
/function="enzyme; Biosynthesis of cofactors, carriers:
Biotin"
/note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
/codon_start=1
/transl_table=11
/product="biotin synthesis, sulfur insertion?"
/protein_id="AAC73862.1"
/db_xref="GI:1786992"
/translation="MAHRPRWTLTQVTELFKEPLDLLFEAQVHRHFDPRQVQVST
LISIKTGACPECKYQPSRYKTGLEARLMEVEQVLESARKAAGSTRFCMGAAW
KNPDRMPYLEOMVQGVKAMGLEACTMTGLTSESQAORLANAGLDYNNHNDTSPF
YGNITTYQERLDLLEKVRDAGIKVCSGGIVGLGETVKDRAGLLQLANLPTPES
VPINMLVKVGTPLADNDVDVAFDPIRTIATVIMPTISYVRLSAGREOMNEQOAMC
FMAGANSIFYGCKLLTTPNPEEDKDLQFLKGLNLPQQTAVLAGDNEQOQKLEALMT
PDTDEYNAAL"
4530..5684
/gene="bioF"
/note="b0776"
4530..5684
/gene="bioF"
/EC_number="2.3.1.47"
/function="enzyme; Biosynthesis of cofactors, carriers:
Biotin"
/note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
/codon_start=1
/transl_table=11
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAC73863.1"
/db_xref="GI:1786993"
/translation="MSWQEKINAAIDARRAALRRRYPVQAGRWLVADROYLNF
SNDVLGLSHHPQIIRAMQCGEORIGSGSGHVSGYVVHQALAEELAEWLIGY SRA
LFIISGFANQAVIAMMAKEDRIADRLSRASLLEASLSPSQLRFAHNDVTHLAR
LIASPCPGQMVVTEGVFSMDGSDSAPLAEIQOVTQOHNGWLMVDVDAHGTGVIGGGRG
SCWLQKVPPELLVTVTFRGFGVSGAAVLCSSTVADYLIQFARHLIYSTSMPPAQAAL
```

Query Match

100.0%; Score 25; DB 1; Length 11022;

```
Best Local Similarity 100.0%; Pred. No. 0.28; BCT 31-JUL-2000
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatctgctgtttgaagcgagcag 25
|||||
Db 3555 GGATCTGCTGTTTGAAGCGAGCAG 3579

RESULT 13
AE004192 12891 bp DNA BCT
LOCUS Vibrio cholerae chromosome I, section 100 of 251 of the complete
DEFINITION chromosome
ACCESSION AE004192 AE003852
VERSION AE004192.1 GI:9655581
KEYWORDS
SOURCE
ORGANISM Vibrio cholerae
Bacteria; Proteobacteria; gamma subdivision: Vibrionaceae; Vibrio.
REFERENCE 1 (bases 1 to 12891)
AUTHORS Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,
Nierman, W.C. and White, O.
TITLE DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae
JOURNAL Nature 406 (6795), 477-483 (2000)
MEDLINE 20406833
REFERENCE 2 (bases 1 to 12891)
AUTHORS Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I.,
Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,
Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,
Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1..12891
/organism="Vibrio cholerae"
/strain="N16961"
/serotype="O1"
/db_xref="taxon:666"
/chromosome="I"
/note="biotype: El Tor"
complement(84..1370)
/gene="VC111"
complement(84..1370)
/gene="VC111"
/note="VC111"
/note="similar to GB:J04423 SP:P12995 PID:457106
PID:490217 GB:U00096; identified by sequence similarity:
putative"
/codon_start=1
/transl_table=11
/product="adenosylmethionine-8-amino-7-oxononanoate
aminotransferase"
/protein_id="AAF94270.1"
/db_xref="GI:9655582"
/translation="MMMDLAFDRCHIWHPTSTLPLTCYVPSRADGVMTLTDGRSL
VDCSSWAAIHYGNPOLNAAMKSOIDAMSHVMEFGIITHAPAIELCRKLVAMTPQ
QVFLADSGVAEVAMKALQYWOAKGEARQRELTFRNGYHGDTFGAMSVCDPDS
MHSIKCYLLENIPAPAPQSGWDEMDRDMVGFARLMAHREHEIAAVIEPIVQAGG
MRYHPWLKRIKICDREGILLIADEIATGRTGKLFACERAEIADPILCLGKALT
GGMTLSATLTITREVAETISNGEAGCFMHGPTFMGNPLACAAASLAIESQDWOQ
VANAEAFBQOLPKLNDSPRVKOTRWLCAIGVYVTHLPVNNNEVITQALFVHGWIIRPF
GRLLYMPPIYQEPHQLKLVDAAMAALNQDPCFKQ"
1497..2549
/gene="VC1112"
1497..2549
```



```

RESULT 14
LOCUS AE005258
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82 of 155.
ACCESSION AE005258
VERSION AE005258.1
KEYWORDS GI:12513751
SOURCE Escherichia coli O157:H7 EDL933.
ORGANISM Escherichia coli O157:H7 EDL933
REFERENCE 1 (bases 1 to 13501)
AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,I., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A., and Blattner,F.R.
TITLE Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
JOURNAL Nature 409 (6819), 529-533 (2001)
MEDLINE 21074935
PUBMED 11206551
REFERENCE 2 (bases 1 to 13501)
AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,I., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A., and Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source
1..13501
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/feature="enterohaemorrhagic"
<1..7576
/misc_feature
/feature="O-island #36; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655; Cryptic prophage CP-933K; includes one copy of the 13 bp direct repeat that flanks the prophage"
66..665
/gene="lomk"
/feature="lomk"
66..665
/gene="lomk"
/function="putative membrane; Other or unknown (Phage or Prophage Related)"
/feature="Residues 1 to 199 of 199 are 72.86 pct identical to residues 1 to 199 of 199 from GenPept 118 : gi17532789|gb|AAFG6231.1|AF151091_2 (AF151091) Lom [prophage P-ElbA]"
/codon_start=1
/transl_table=11
/product="putative outer membrane protein of prophage CP-933K"
/protein_id="AAG55137.1"
/db_xref="GI:12513752"
/translation="MRKVCAAIILSAICLAVSGVPWASEHQSTLSAGYLAHSTDAPEG SDLLGKINVKYEFDTLGLTFSYANAEDQKTHYSDTRWHEDYVRRNRFVSMAG PSVRVNEFSAVAMAGVAYSRVSTFSGDYFRVTDNKRKTHDVLTGSDDDARYSNTSLAW GAGVQNFPTESVAVDVAYEXSGSDPRTDGFIVGVGYKF"
719..2041
/gene="20982"
/feature="20982"
/function="putative structure; Structural component (Phage or Prophage Related)"
/feature="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/translation="MKTPISQSNFNSAPAGYSAPTAPNRAENAYADYVLDIGKRPL SAADLSNVYESVIRAVHDSRLDIOHTVDMIGNTVLDLSRSQTPFDAYSYGHNK VHGICIKRYEVELNEESVKIDDQSLTNCNEIYEDVGQEPFPFICEAGYENEEPY VSFSAVPTDSYEMPSWQGLIHEI IHVVTGSSDPGSDSIELGPTLILARRVAQELG WSPDFKGYAFPEREAHLRLNLALRQAAMRHENERAFFERLGTISDRYASPDPT

```

```

EYSAVSNIGYGFIOQHDPPLAINDNLQDANQIQYHGAPYIFTFTGVDWKHNOQ"
5350..6231
/gene="20989"
CDS
5350..6231
/gene="20989"
/function="orf; Other or unknown (Phase or Prophage
Related)"
/notes="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/db_xref="GI:12513757"
/translation="MLSPYSVNLGCSWNSLTNLTSPDNRLVSSVRDAAVHSDNGAOV
KVGNTYRVATDNKFTRESHGCFNLHLRUGWPKGEISRKIEVLMNAPVSAAM
ERGIVHSNPDLPPVDYAPPELPSVYNRLSPGVNIGKGNVAVVEDADAIVKLKM
FTTSQNEEVTSEVRCFNQYAGSAEKIYVNGNDIIGIRMDKINGSNLNLSLPAQ
AEHALYDMFDRLKQGLFVDTETNVLIDRAKNFNPDISSYNSDRSWSSEQIMQ
SYHGKQDLISVLSKI"
6462..7160
/gene="20990"
CDS
6462..7160
/gene="20990"
/function="orf; Other or unknown (Phase or Prophage
Related)"
/notes="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/db_xref="GI:12513758"
/translation="MRPTSLNVLHQSSRSSMSDDTIESLVKASSVQVTKNNPQLRF
QDHNHYIQEALDGKSTGTGTVLLNLAIESIRLKTETVVIHLSNRLGVMYHARD
IGAHNRGTGSGFNNAVETPCGEGISVDFHATVFEHLHFHNLNGERLKVES
SRASQRYSPLLLEARTVGLGAFSEVLSNKFHEEIGMPRTSYPPXDSALHTDDNT
VSLGFOQVRLHPL"
complement(7668..8144)
/gene="ybhB"
CDS
complement(7668..8144)
/notes="20992"
/gene="ybhB"
/function="orf; Unknown function"
/notes="Residues 1 to 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGLKPHRVFNGMGYDGDNISPLAWDDVPACTK
SFVVTCDPDAPTCGSHWHVYVNLPAATRVLPQGGSLVAPDGVLTQRTDFGTG
YDGAAPKPGETHRIYFIVFHALDVERIDVDEGASGMVGFNVHPSLASASITAMFS"
complement(8203..9492)
/gene="bioA"
CDS
complement(8203..9492)
/notes="20993"
complement(8203..9492)

Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatctgctgttgaagcgcagcaq 25
|||||
Db 9641 GGATCTGCTGTTGAAGCGCAGCAG 9665

RESULT 15
AP002553
LOCUS AP002553 297816 bp DNA BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION AP002553 BA000007
VERSION AP002553.1 GI:13360211

```

KEYWORDS  
SOURCE

## ORGANISM

REFERENCE  
AUTHORS

## TITLE

JOURNAL  
MEDLINEREFERENCE  
AUTHORS

## TITLE

JOURNAL  
MEDLINEREFERENCE  
AUTHORS

## TITLE

JOURNAL  
MEDLINEREFERENCE  
AUTHORS

## TITLE

JOURNAL  
MEDLINEREFERENCE  
AUTHORSTITLE  
JOURNALCOMMENT  
FEATURES

## Source

## gene

## CDS

```

Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RMD 0509952)
DNA.
Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (sites)
Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S.,
Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T.,
Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
Sasakawa, C. and Shinagawa, H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780
2 (sites)
Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,
Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and
Hayashi, T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
3 (sites)
Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S.,
Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and
Shinagawa, H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
4 (sites)
Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,
Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C.,
Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and
Shinagawa, H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
5 (bases 1 to 297816)
Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and
Hayashi, T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
Fax:81-6-6879-2047)
genome project.
Location/Qualifiers
1..297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RMD 0509952"
/db_xref="taxon:83334"
79..1245
/gene="ECs0753"
79..1245
/gene="ECs0753"
/notes="similar
100 in 388 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="succinyl-CoA synthetase beta subunit"
/protein_id="BAB34176.1"
/db_xref="GI:13360212"

```



```

/translacion="MNLHEYOAKOLFARYGLPAPVGVYACTTPREAEAAASKIGAGPW
VKOVHAGGKAGKGVVNSKEDIRAFENWLGKRLVYTYQTDANGOPVNIIVERAAT
DIARELIGAVDRSRVPMSTEGGVEIEKVAETPLHILKHLVDPLDTPMPYOG
RELAKLGLEKLVQQTKEIPMGGLATIFLERDLALIEINPLVITKQGLDCLDKLGA
DGNALFROPDLREMRDQSGPREAQAQWELNYVALDGNIGWNGAGLAMGTMDIV
KLHGEPANFLDVGCGATKEVTEAFKILSDDKKAVLVNIFGIVRCDLIADGIIG
AAVGVNVPVVVRELGNAELGAKKGLADSLNIIAAGLTDAAQVVAAVEGK"
1245..2114
/gene="ECS0754"
CDS
1245..2114
/gene="ECS0754"
/notes="similar to SUCD_ECOLI gi|1786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/translacion="MSILDKNTKVICQPTGSGTSHSQAIAYGTMVGGVTPGKG
GTHLGLPVFNTVEAUAATGATASVITVPAPFKDSILEDAIDAGIKLITITTEGPT
LDMLTVKLDAGVGMIPNCPGVITPGECKIGIQGHHPKGVKGLIVSRSGTLTYE
AVKQTDYFGQSCVIGIGDPIPGSNFIDILEMFKDPQTEAIVMIGEIGSGAEEA
AAIYKEHTVKPVVGVLGVTAPKGRMGAGAIAGGKTADERFAALEAAGVKTVERS
LADIGEALTKLK"
2781..3686
/gene="ECS0755"
CDS
2781..3686
/gene="ECS0755"
/notes="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcriptional activator (gcv operon activator) -
Escherichia coli gi|417043|sp|P32064|GCVA_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/translacion="MRGKIPKTELLVTEFVVARHESYTRAAEELALTSQAVRQVSAL
BEFLTPFLFSKRIKIFUNDGKYVIGVETLNLKLRDNTIMTQPTVOVIELAVN
PFSTHMLPNLHEFTKLHPDIIVNIHSLANGDPLNREYDAVIMRENFCAPAVEY
LFEELPVCSSLLANSQKLSVAELLTEPLHQSITRGWEWFALSVSSPLVN
NGPREDLLSMLAAVRSNLGVALLRFAIQHDLSDGDMVPCDVFIRGNRFIMTWQE
EKSDSPHLQKQFREWLLAKSVVPQM"
complement(3720..4322)
/gene="ECS0756"
CDS
complement(3720..4322)
/gene="ECS0756"
/notes="probable cob(I)alamin adenosyltransferase, similar
to cob(I)alamin adenosyltransferases (corrinoid
adenosyltransferases) e.g. [Escherichia coli]
gi|115148|sp|P13040|BTUR_ECOLI percent identity 67 in 200
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative cob(I)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translacion="MEARLSTERHVROOKLKEOVDTRVAAIEKKGILLIVFTGNGKG
KSTAPGTTRAVAGHKTGVGAQYKNGQMDNGEYNLLQPLGVFEHMGTFWTQNR
QADIDAAKVESKRLADRYDLVLDLTYMLAYHLDTEEVIAISLQNRPAQGSV
IVTGRCHSQILKMDTVSEIRPVKHFNDNGIQAPGIDW"
complement(4332..5984)
/gene="ECS0757"
CDS
complement(4332..5984)
/gene="ECS0757"
/notes="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarase) - Escherichia coli
gi|120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545

```

```

aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"
/translacion="MSKPFILQEPFLQNKDTEYTLISDQHTVTVELDGEVVIKIAPE
ALTLARQAFYASFLRSALHQVAAIILDDQASNDKYVALQTLRNVEVSAKVLP
NCPDGTATIIAVSKGQOIWNTGNDAAEALSKGIYTFQENNLFSQNALPDMYTEVNTQ
TNLPAQIDISAVAGDEHYELCVNKGSGSANKAALYOETKSLIQPEKLFLEKMKSL
EASQPPYIFAFVYVGLSADOTLTKIAKASTKYDNLPTSGNEQGGQAFRIDELEKVL
EASQPGIQAQFGKYFAHDRIVRILPRHGGSCPIAMALSCSADRNIKAKINKHGIWL
EKLEHNPQGOVIPASLEENHQAQVOLDNRPLRDVMDLARIPVGTFRVLSGPIVVAR
DIAHAKIKARLDSGPEMPEYKHHIIVYAGPAKTPENMACGSLGPTTGGRMGDIYDF
OAGGSLVMLSKGNRSQOYTDACHKHGGNLSGICGAALLAQEYVYKSLRCLEYPBLG
MEAVWMVEVENLPAFLVLDKDGNNFFSQFEQQRCAACSPAGH"
complement(6092..7372)
/gene="ECS0758"
CDS
complement(6092..7372)
/gene="ECS0758"
/notes="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
gi|121467|sp|P24943|GLTT_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transpor"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translacion="MKKISLTMLLALVLGMLIGVNLNTASPTAKLYAQBISIFT
TIFRLRLKIIAPLVSTLVVGIKMGDAKALGRIFSKTLFLFCASLSIALGLITY
NFPMPGTGINFVAHGAETTCVAAEPFLKVFISHAFPTSIVDMAHNEILQIVVFSI
FLAGCSITAIAGEKSAIVHALDSLAAHMLKLTGYVFLAPLTFVFAALISALIEGLAVM
VSAGIFMGEEFYTMILLIIGLAIIVYVGPICRRLTRALSEPALLAFTTSSEAAFP
GTLEKLPQGVSPKIASFVLPVIGYFNLVGSMAYCSFATVFIQAACNHLISGEQITM
LLIILMTSKMGAGVPRASVMVIAATLNQFNIPEAGLILLMGVDPFLDMGRSATNMVN
AMGAAMVSRGEHGEHFGEGCRGKALKPNESNALP"
complement(7533..7850)
/gene="ECS0759"
CDS
complement(7533..7850)
/gene="ECS0759"
/notes="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"
Query Match 100.0%; Score 25; DB 2; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 gqatctgctgtttgaagcgacag 25
|||||
Db 92373 GGATCTGCTGTTTGAAGCGCACGAG 92397
|||||
RESULT 16
LOCUS E00893 1121 bp DNA PAT 29-SEP-1997
DEFINITION Genomic DNA encoding biotin Synthetase.
ACCESSION E00893
VERSION E00893.1 GI:2169154
KEYWORDS JP 1986149091-A/1.
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 1121)
AUTHORS Hirono.Y., Kojima,T. and Kimura,H.

```

```

TITLE      DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
JOURNAL    PRODUCTION OF BIOTIN
            Patent: JP 1986149091-A 1 07-JUL-1986;
            NIPPON SODA CO LTD
COMMENT     OS Escherichia coli
            PN JP 1986149091-A/1
            PD 07-JUL-1986
            PF 24-DEC-1984 JP 1984272605
            PI HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
            C12N15/00.C12N13/18, C12P13/18, (C12N1/20, C12R1/19), (C12P13/18, PC
            C12R1/19);
            CC strandedness: Double;
            CC topology: Linear;
            CC hypothetical: No;
            CC anti-sense: No;
            CC *source: strain=Escherichia coli Nsl01;
            CC Feature is identified by experimental;
            PH key Location/Qualifiers
            FT CDS 42..1079
            FT     /product='biotin synthetase'.
            FT     Location/Qualifiers
            FT     1..1121
            FT     /organism='Escherichia coli'
            FT     /db_xref='taxon:562'
BASE COUNT 289 a 296 c 325 g 211 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 10; Length 1121;
Best Local Similarity 96.0%; Pred. No. 1.4;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ggatctgctgttgaagcgcagcag 25
Db 104 GGATCGCTGTTGAAGCCGACGAC 128

RESULT 17
AF248314
LOCUS      8227 bp DNA BCT 24-JAN-2001
DEFINITION Uncultured bacterium pCosAS1 urocanase-like protein (hutU) gene,
            partial cds; histidine ammonia-lyase-like protein (huth), DAPA
            aminotransferase BioA (bioA), biotin synthase BioB (bioB), 7-KAPA
            synthetase (bioF), biotin biosynthesis Bioc-like protein (bioc),
            and dehydrobiotin synthase Biob (biob) genes, complete cds; ABC
            transporter-like protein (elsa) gene, partial cds; and unknown
            gene.
ACCESSION AF248314
VERSION    AF248314.1 GI:12407610
SOURCE     uncultured bacterium pCosAS1.
            uncultured bacterium pCosAS1.
            Bacteria; environmental samples.
            1 (bases 1 to 8227)
REFERENCE  Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streit, W.R.
            Direct Cloning from Enrichment Cultures, a Reliable Strategy for
            Isolation of Complete Operons and Genes from Microbial Consortia
            Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
JOURNAL    1113432
PUBMED     2 (bases 1 to 8227)
REFERENCE  Entcheva, P., Liebl, W. and Streit, W.R.
            Direct Submission
            Submitted (21-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Griesbachstr. 8, Goettingen 37077, Germany
FEATURES   Location/Qualifiers
            1..8227
            /organism='uncultured bacterium pCosAS1'
            /db_xref='taxon:143793'
            /clone='pCosAS1'
            /note='unknown organism, cosmid clone derived from
            environmental consortium'
            <1..502
            gene

```

```

            /gene="hutU"
            <1..502
            /gene="hutU"
            /note="putative urocanate hydratase"
            /codon_start=2
            /transl_table=11
            /product="urocanase-like protein"
            /protein_id="AAG53593.1"
            /db_xref="GI:12407618"
            /translation="PRETISFQGLPAPICWVGLWQKLGAFNEMVRRGEVSAPIVI
            GRDLDSGVSASPNRETEAMQDSDAVSDWPLLNALLNTASGATWYSLHHGGVGMGF
            SQHAGMVICDGTDDAAARTARVLHNDPATGVNMRHADAGYDIAIDCAKEQGLNLPMA
            ATOGQH"
            502..2031
            /gene="hutU"
            502..2031
            /gene="hutU"
            /note="putative"
            /codon_start=1
            /transl_table=11
            /product="histidine ammonia-lyase-like protein"
            /protein_id="AAG53586.1"
            /db_xref="GI:12407611"
            /translation="MNALLTPTGLTLAQLRWVWQPLQLTLDESAREAINDSVACVE
            AIVAGRTAYGINTGFGLLAQATRIATHLENLQRLSLVSHAAGVGPIDDDIVRLMMV
            LKINSLARGFSGIRLSVIOALIALVNAVYSVDPAKGSVSGSDLAFLAHMSILTLGE
            GKARYGEWLPAATALQKAGLAPVTLAAKEGLALLNGTOASTAFALRGLFEADLPAS
            AVVCGALTTEAVLGSRPPDARIHEVRGQGDAAALFRHLVTLDTSAIESHHNCDK
            VQDYSRLCQPMGACLTMOMROVAEVLIVSNVSDNPLVFAAENEMVFRGNFAEP
            VAMAADNLALATAETALSERIALMMDKHMSQLPPLVNRNGVNSGFMIAQVTAAL
            ASNKGCLHPHTTSVDKPCRPASQEDHVSMAFPAAGRRLWEMAGNTRGLVAVELAAAC
            QGADLRDLGTLSSPLPQARQSGCEQVAHYDDDRFFAPDTEAAISLLNKGSLVGLLPAP
            L"
            complement(2130..2606)
            /note="ORF1; similar to hypothetical 17.1 kDa protein in
            modC-bioA intergenic region"
            /codon_start=1
            /transl_table=11
            /product="unknown"
            /protein_id="AAG53587.1"
            /db_xref="GI:12407612"
            /translation="MKLISQDLRDGDLRLRVFNGMGVEGDNISPHLAWDEVPSGTK
            SFVTCYDPAPTDSGWHWIVANIPADTRVLPQSGSSVALPEGAVETRTDFKAG
            YGGAAPPKGETHYIFTVHALNVERIEVDEASGAMGVNFHFTLGSASITAMYS"
            complement(2651..3940)
            /gene="bioA"
            complement(2651..3940)
            /gene="bioA"
            /note="7,8-diamino-pelargonic acid aminotransferase"
            /codon_start=1
            /transl_table=11
            /product="DAPA aminotransferase BioA"
            /protein_id="AAG53588.1"
            /db_xref="GI:12407613"
            /translation="MTDLDLADROHIIHPTYSMTQPLPVPVVAHGCGLTASGEQ
            LVDMSSWMAALHGYNHPLNAAKQAQIDRMHVMFGGITHQPAVDLCRLVAMTPEP
            LECVPAPSGSVAVAKMALQYWHAKESQRFITFNGYHGTFGAMSVCDDPNS
            HSLMKGLYPENLAPAPQSRFDGAWDEMVGVLAPLMAAHRHEIAAIVLEP IVOGAG
            GMRVYHPWLKIRRMCDREGLLLIADETATGFGTKLFACEHAGITPDILCLGKAL
            TGGTMTSATLTTRQVETISNCEAGCFMHGPTFMGNPLACAVASESLALESERQD
            OVAALIESLOELAVARDESEVADRVVLGAGVETTHPVNMAALQREFVARGVWRP
            FGKLIYLPPIVIRPDQLSKLITRAVVDAVQAPAHFTR"
            4029..5069
            /gene="bioB"
            4029..5069
            /gene="bioB"
            /codon_start=1
            /transl_table=11
            /product="biotin synthase BioB"
            /protein_id="AAG53589.1"
            /db_xref="GI:12407614"
            /translation="MAHQARWTLTSQVATLAFPEKPLLELLEFAQTHROHFDPOQIOVST
            LLSIKTGACPEDCKYCPQSAKYARKTGLESERLMEVEQVLESARQAKNAGSTRFCGAAW

```

KNPHERDMPYLEOMVGVKAMGLQACMTLGLTDDHQOARLASAGLDYNNHNLTFPEF  
 YGNITRTYQERLDTLKVVDAGIKVCSGIVGLCETVTDTRAGLLQLANLTPPPES  
 VPINMLVKVGTPLADNDVYAFDXIRTSVAXVMXPTXILRLSTGREQNEQTCAMC  
 PMAGANSIFYGCKLLTTPNPEEDKQVLFRLGLNPOQTAVLTGDNESQOHLQOOLFEN  
 ADTDOFYNAATV"  
 5066..6223  
 /gene="biof"  
 5066..6223  
 /gene="biof"  
 /note="7-keto-8-amino-pelargonic acid synthetase; similar  
 to Biof"  
 /codon\_start=1  
 /transl\_table=11  
 /product="7-KAPA synthetase"  
 /protein\_id="AAG53590.1"  
 /db\_xref="GI:12407615"  
 /translation="MTWQORIDAALDARRAADALRVTPVENGAGRELIRORRYCNF  
 SGNDYLGSGHPEIIRAQOAGQGVGSGGSHVGYTHAQALEHLAGWLYCPRA  
 LLFTSGFAANQALTAALTDKEDRIVADRLSHASLLEASHSPAOLRRFAINDVAOLDS  
 LLDKPCHQQLVTEGIFSMGDSAPGLVHAQAQRQGSWLLVYDAHGIGVTPGEGRG  
 SAASGVVPELLVVTFGKGVGSAALVCSDAVADYLLQFAHLLIYSTMPAQAVAL  
 LAAFNVRISDEDNARRQALALIDQFRAGVRDLPVQITASQSAOTPLIVGENARALHL  
 AQLRERGIWTVAIRPTVPAGTARLRLTLTAHQPADITALVEALHVAGE"  
 6207..6962  
 /gene="bioc"  
 6207..6962  
 /gene="bioc"  
 /codon\_start=1  
 /transl\_table=11  
 /product="biotin biosynthesis BioC-like protein"  
 /protein\_id="AAG53591.1"  
 /db\_xref="GI:12407616"  
 /translation="MSQVKNQAIAAAFGRAGQSYSRHDALQRQADGLRALLAENTFA  
 SYLDACGPGNSLAWROTGSVVTAMDLCAPMLEEARNQAADSYLQADIAEIPLADA  
 QPSLAWSLAVQXCASLAPRALAEMLRWTLQGGKVAFTLLNGSLPELNAQWQAVDTOP  
 HANRLSGQOVNDLSGWQYRCVQTVTLTFDDLSAMHSUKGICATHLHAGRAAKPL  
 TRGQLQALQAQAPQAGKPEPLSYQLFHGIIRD"  
 6940..7680  
 /gene="bioc"  
 6940..7680  
 /gene="bioc"  
 /codon\_start=1  
 /transl\_table=11  
 /product="dethiobiotin synthase BioD"  
 /protein\_id="AAG53592.1"  
 /db\_xref="GI:12407617"  
 /translation="MELSNVTEREYVTGDTVEGKTAVSTALQQAALGRHTAGYKP  
 VASGEMTAQGRNSDALALQNSLSALSYPAVNPYTPAEPSTPHIVSADEQPIEFA  
 TISAGLRALETDQADVLVEGAGGWFTPLSQEOTFADWMAERLPVILVVGKLGCIINH  
 AMLTQAIRQAGLHFGAGVANGVYVPGKRAEYMATLRVLSAPLGEIPWLGQADT  
 SAVGQLDLTALTSPSSGQAGHPAR"  
 complement(7619..>8227)  
 /gene="elsa"  
 complement(7619..>8227)  
 /gene="elsa"  
 /note="putative urea/short-chain amide ABC transporter"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ABC transporter-like protein"  
 /protein\_id="AAG53594.1"  
 /db\_xref="GI:12407619"  
 /translation="DCXLRTRGXENXLYSCIAGLLPIASGTMVWHEAPGAVNOAGTEI  
 XYVQDRIIFSLQTVENLHTAQMAKGIAGSLSRDVEFFFPFLPLKHLRASGLND  
 QOYLALNFTPLRLJLIDEPSTGQFTVQLKALLRLKRGSLVLLAEQHL  
 LIRVRADRECLLYRGRNVAQGRVSELDPLIAHWGPTASQ"  
 1512 a 2642 c 2605 g 1453 t 15 others

BASE COUNT 1512 a 2642 c 2605 g 1453 t 15 others  
 ORIGIN

Query Match 89.6%; Score 22.4; DB 2; Length 8227;

Best Local Similarity 95.8%; Pred. No. 4.9;

Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gatctgctgtttgaagcgcagcag 25  
 |||||  
 Db 4092 GAATGCTGTTGAAGCGCAGCAG 4115  
 |||||  
 RESULT 18  
 ECOTIOB 128 bp DNA BCT 03-JAN-1995  
 LOCUS Escherichia coli biotin (bioB) gene, early terminator region.  
 DEFINITION M27731  
 ACCESSION M27731.1 GI:341755  
 VERSION  
 KEYWORDS bioB gene; biotin.  
 SOURCE Escherichia coli (strain K-12) DNA.  
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Escherichia.  
 REFERENCE 1 (bases 1 to 128)  
 AUTHORS Nath, S.K.  
 TITLE Attenuation of transcription of biotin genes in Escherichia coli  
 JOURNAL Can. J. Microbiol. 34 (12), 1288-1296 (1988)  
 MEDLINE 89167942 Location/Qualifiers  
 FEATURES  
 source  
 1..128  
 /organism="Escherichia coli"  
 /strain="K-12"  
 /db\_xref="taxon:562"  
 gene  
 4..128  
 /gene="bioB"  
 -10\_signal  
 4..9  
 /gene="bioB"  
 mRNA  
 20..>128  
 /gene="bioB"  
 terminator  
 103..121  
 /gene="bioB"  
 /note="early terminator"  
 BASE COUNT 31 a 28 c 34 g 35 t  
 ORIGIN

Query Match 84.0%; Score 21; DB 2; Length 128;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatctgctgtttgaagcgcagcag 21  
 |||||  
 Db 24 GAATGCTGTTGAAGCGCA 44  
 |||||

RESULT 19  
 AC004313 80993 bp DNA INV 02-APR-1998  
 LOCUS Drosophila melanogaster (Pl DS02368 (D205)) DNA sequence, complete  
 DEFINITION sequence.  
 AC004313  
 AC004313.1 GI:3006211  
 VERSION HTG.  
 KEYWORDS Drosophila melanogaster (Subclones in Sac from Pl clone DS02368  
 SOURCE (D205)) DNA.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 80993)  
 AUTHORS Celisner, S.E., Aghavani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,  
 Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A.,  
 Hummasti, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B.,  
 Lomont, M.A., Mazda, P., Mok, M.S., Moshrefi, A.R., Moshrefi, M.,  
 Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., PUNCH, D., Snir, E.,  
 Twomey, B., Wan, K.H., Whitelaw, K.R., Yee, A., Zhang, R., Zieran, L.L.  
 and Kimmel, B.  
 TITLE Sequencing of Drosophila chromosome 2R, region 57B14-57C1  
 JOURNAL Unpublished (1997)  
 REFERENCE 2 (bases 1 to 80993)



```

pieces.
ACCESSION AC014412
VERSION AC014412.1 GI:6436923
KEYWORDS HTG: HTGS_PHASE2.
SOURCE fruit fly
ORGANISM Drosophila melanogaster

REFERENCE
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10212722 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1. .86446
Location/Qualifiers
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 23099 a 20123 c 20064 g 23160 t
ORIGIN
Query Match 76.8%; Score 19.2; DB 63; Length 86446;
Best Local Similarity 87.5%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gatctgctgtttgaagcgcagcag 25
||||| ||||||| |||||||
Db 45648 GATGCTCTGTTGAAGCTCAGCAG 45625

RESULT 23
AC009213 182525 bp DNA INV 17-FEB-2001
LOCUS Drosophila melanogaster, chromosome 3R, region 98D-98D, BAC clone
DEFINITION BACR09F18, complete sequence.
ACCESSION AC009213
VERSION AC009213.5 GI:12957654
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

REFERENCE
AUTHORS Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Ibezawam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirkas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 3R, region 98D-98D
Unpublished
2 (bases 1 to 182525)
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,

TITLE
JOURNAL
REFERENCE
AUTHORS

```

Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,  
 Svirkas,R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
 Rubin,G.M.  
 Direct Submission  
 Submitted (06-AUG-1999) Drosophila Genome Center, Lawrence Berkeley  
 Laboratory, MS 64-121, Berkeley, CA 94720, USA  
 On Feb 17, 2001 this sequence version replaced gi:6957913.  
 Sequence submitted by:  
 Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory, MS 64-121  
 Berkeley, CA 94720  
 This sequence was assembled using end sequences from a whole genome  
 shotgun and from subclones of this BAC and its neighboring clones.  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our sequence  
 archive web site (<http://www.fruitfly.org/sequence/>) or send email  
 to [bdgpf@fruitfly.berkeley.edu](mailto:bdgpf@fruitfly.berkeley.edu).  
 Location/Qualifiers  
 1. .182525  
 /organism="Drosophila melanogaster"  
 /strain="y; cn bw sp"  
 /db\_xref="taxon:7227"  
 /chromosome="3R"  
 /map="98D-98D"  
 /clone="BACR09F18 (D812)"  
 /clone\_lib="RPCI-98 (Roswell Park Cancer Institute  
 Drosophila melanogaster BAC library, partial ECORI in  
 pBACE3.6)"  
 BASE COUNT 49013 a 41841 c 41897 g 49774 t  
 ORIGIN  
 Query Match 76.8%; Score 19.2; DB 4; Length 182525;  
 Best Local Similarity 87.5%; Pred. No. 2e+02; 3; Indels 0; Gaps 0;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 gatctgctgtttgaagcgcagcag 25  
 ||||| ||||||| |||||||  
 Db 33962 GATGCTCTGTTGAAGCTCAGCAG 33985  
 RESULT 24  
 AC008028 197348 bp DNA INV 17-FEB-2001  
 LOCUS Drosophila melanogaster, chromosome 3R, region 98C-98D, BAC clone  
 DEFINITION BACR44L13, complete sequence.  
 ACCESSION AC008028  
 VERSION AC008028.3 GI:12957647  
 KEYWORDS HTG.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 197348)  
 Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,  
 Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,  
 Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,  
 Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,  
 Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,  
 Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,  
 Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,  
 Ibezawam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A.,  
 McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,  
 Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,  
 Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,  
 Stapleton,M., Strong,R., Svirkas,R., Tector,C., Williams,S.M.,  
 Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.  
 Sequencing of Drosophila chromosome 3R, region 98C-98D  
 Unpublished  
 2 (bases 1 to 197348)  
 Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
 Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
 Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
 Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
 Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,  
 Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,

Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Svirska, R.R., Wan, K.H., Weinburg, F., Zhang, R., Zieran, L.L. and Rubin, G.M.

Direct Submission  
Submitted (12-JUL-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Feb 17, 2001 this sequence version replaced gi:5670462.  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [bdgpf@fruitfly.berkeley.edu](mailto:bdgpf@fruitfly.berkeley.edu).

# FEATURES

source

```
1. 197348
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="98C-98D"
/clone="BACR14L13 (D789)"
/clone.lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBACe3.6)"
```

BASE COUNT 54902 a 42954 c 44141 g 55351 t

## ORIGIN

Query Match 76.8%; Score 19.2; DB 4; Length 197348;  
Best Local Similarity 87.5%; Pred. No. 2e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcgcagcag 25

||||| ||||| ||||| ||||| |||||  
Db 188662 GATGTCCTGTTGAAGCTCAGCAG 188685

## RESULT 25

AE003765

LOCUS

DEFINITION

AE003765 AE002708

VERSION

KEYWORDS

HTG.

SOURCE

ORGANISM

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 239171)

Adams, M.D., Celisner, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,

Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,

George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,

Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,

Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D.,

Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor

Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J.,

Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A.,

Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y.,

Benos, P.V., Bertram, B.P., Bhandari, D., Bolshakov, S., Borkova, D.,

Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C.,

Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I.,

Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de

Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M.,

Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C.,  
Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferreira, S.,  
Fleischmann, W., Foslter, C., Gabrielian, A.E., Garg, N.S.,  
Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z.,  
Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J.,  
Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,  
Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,  
Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,  
Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A.,  
Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C.,  
McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Moberly, C.,  
Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,  
Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,  
Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S.,  
Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K.,  
Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Slden-Klamos, I.,  
Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C.,  
Stapleton, M., Strong, R., Sun, E., Svirska, R., Tector, C., Turner, R.,  
Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A.,  
Weinstock, G.M., Weissbach, J., Williams, S.M., Woodage, T.,  
Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F.,  
Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H.,  
Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O.,  
Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.  
The genome sequence of *Drosophila melanogaster*  
Science 287 (5461), 2185-2195 (2000)  
20196006  
2 (bases 1 to 239171)  
Adams, M.D., Celisner, S.E., Gibbs, R.A., Rubin, G.M. and Venter, J.C.  
Direct Submission  
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
On Oct 9, 2000 this sequence version replaced gi:7301668.  
Location/Qualifiers  
1. 239171  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/chromosome="3R"  
complement(<3698. .>4543)  
/gene="CG14065"  
/product="CT3634"  
/db\_xref="FLYBASE:FBan0014065"  
/db\_xref="FlyBase:FBgn0039579"  
/evidence=not\_experimental  
complement(<3698. .>4543)  
/gene="CG14065"  
/map="98C3-98C3"  
/db\_xref="FLYBASE:FBan0014065"  
/db\_xref="FlyBase:FBgn0039579"  
/evidence=not\_experimental  
complement(3698. .4543)  
/gene="CG14065"  
/note="CG14065 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0014065"  
/db\_xref="FlyBase:FBgn0039579"  
/evidence=not\_experimental  
/protein\_id="AAF56784.1"  
/db\_xref="GI:7301669"  
/translation="MSSKEETPTSGGAAPVATVAAGSATTSYANVYONLETKKSGTS  
NTTTTATATVDNKENQNLKFSKSWSEAAAAAGSPVGSVSLAQTEKRAAAS  
GCDNTAENSSELDNDNNFVPLSHRRDRKRAKREKPDQPPAGGQDGPQAGGR  
RVPGSSTSVSRPRAPRSPRRKSAAGAGNAGAVSSAVPGGPKVHKDKDT  
SPSASIEAGSGSGNEAKSADGDDQAGEKGAAGAAPPKRFRTAAPPKPVNWKVSDQMI  
ESLSD"  
complement(join(7319. .7693,7896. .>9739))  
/gene="CG1345"  
/product="CT3038"  
/db\_xref="FLYBASE:FBan0001345"  
/db\_xref="FlyBase:FBgn0039580"  
/evidence=not\_experimental  
complement(<7319. .>9739)  
/gene="CG1345"







affects morphogenesis and is a suppressor of position-effect variegation

JOURNAL Mech. Dev. 96 (1), 67-78 (2000)

PUBMED 10940625

REFERENCE 2 (bases 1 to 16147)

AUTHORS Stroedicke, M.

TITLE Direct Submission

JOURNAL Submitted (15-JUL-1999) Stroedicke M., Freie Universitaet Berlin, Institut fuer Genetik, Arnimallee 7, 14195 Berlin, GERMANY

FEATURES Location/Qualifiers

1..16147

/organism="Drosophila melanogaster"

/strain="Karsnaes"

/db\_xref="taxon:7227"

/chromosome="3R"

/map="86B1-2"

Join(3888..4290,11666..13465,13526..15207)

/gene="Domina"

3888..15207

/gene="Domina"

3888..4290

/gene="Domina"

/number=1

Join(4231..4290,11666..13465,13526..13825)

/gene="Domina"

/function="transcription factor/chromatin modifier"

/codon\_start=1

/product="DOMINA protein (WHN-homologue)"

/protein\_id="CAB51032.1"

/db\_xref="GI:5531247"

/db\_xref="FLYBASE:FBgn0015660"

/db\_xref="SPTREMBL:O9XTP7"

/translation="MFELEDYSSGTHGEFFSKYADAAGPSLDFFVSDSMQMLNVDIR

AEIANVGGSSDTSLSLDTLEALISANNOSNGSSQASYNANANFLTSSGLHAS  
PTAKWGSANFNSDYADLGACVNPISYMLNTNSTSAGFPKPKNKKTASSTQGRS  
GAVPSPAERDQKSHLTFSPAQKVSAGSMRQDVNAHLPKQISVYVGTGTGTAPAT  
MATVLSGRNRSADVADVKDLVTLERKAQSPVPNSLEELGKGGKSTLLNASVGATN  
TLKAPGIGGLTFANSAAYOKLKOTSLVKSPPGSLSPGAGSNMGLKREDNKRGLQAST  
TPKSTASANSPHQMQNSYLSGSPSSLSSSASPLGNSVNLVNNNTNSGAGSLG  
VKPLQOKVLPVGPFPKPAYISCLIALAKNSRAGSLPVEIISFLCOHFPYFEN  
APSGWNSVRHNLNLFKFEKIERPATNGKRCRWAMNPDRINKMDEEYQKWSRKD  
PAAIRGAMYIPOHLESFERGEMKHSADSDVELDSQSEIESDSEIEHEFEDTWDAM  
LVEEDEEDGDDQIINDFDAEDERHANGNOANLPINHPLLGKSNDFDIEVGDIL  
VDAIDEDDKESVRILINDOQHIIELNADLNATDGYNQOPALKRVRVDINVAIGPAG  
ELEQYIGQRVKVQVQIQQQHPPTTNRMRMLVNRVI"

4291..11665

/gene="Domina"

/number=1

11666..13465

/gene="Domina"

/number=2

13466..13525

/gene="Domina"

/number=2

13526..15207

/gene="Domina"

/number=3

4837 a 3381 c 3364 g 4534 t 31 others

BASE COUNT

ORIGIN

Query Match 75.2%; Score 18.8; DB 6; Length 16147;

Best Local Similarity 90.9%; Pred. No. 2.7e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 gatctgctgttgagcgagc 23

||||||| |||||||

Db 1150 GATCTGCTCTTGAAGCGCAGC 1129

RESULT 29

AC008357/c

LOCUS

AC008357 190673 bp DNA INV 10-MAR-2001

DEFINITION

Drosophila melanogaster, chromosome 3R, region 86A-86A, BAC clone

BACR03L12, complete sequence.

ACCESSION AC008357

VERSION AC008357.7 GI:13270564

KEYWORDS fruit fly.

SOURCE Drosophila melanogaster

ORGANISM Drosophila melanogaster

REFERENCE 1 (bases 1 to 190673)

AUTHORS Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H.,

Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,

Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A.,

Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M.,

Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D.,

Ferrera, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,

Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,

Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,

McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,

Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,

Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,

Shapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M.,

Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Sequencing of Drosophila chromosome 3R, region 86A-86A

Unpublished

2 (bases 1 to 190673)

Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,

Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,

Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,

Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,

Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,

Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,

Pfeiffer, B., Poon, L., Sequiera, A., Sethi, H., Shiri, E.,

Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and

Rubin, G.M.

Direct Submission

Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley

Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Mar 10, 2001 this sequence version replaced gi:7248927.

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome

shotgun and from subclones of this BAC and its neighboring clones.

For further information about this sequence, including its location

and relationship to other sequences, please visit our sequence

archive Web site (<http://www.fruitfly.org/sequence/>) or send email

to [bdgp@fruitfly.berkeley.edu](mailto:bdgp@fruitfly.berkeley.edu).

Location/Qualifiers

1..190673

/organism="Drosophila melanogaster"

/strain="y; cn bw sp"

/db\_xref="taxon:7227"

/chromosome="3R"

/map="86A-86A"

/clone="BACR03L12 (D971)"

/clone\_lib="RPCI-98 (Roswell Park Cancer Institute

Drosophila melanogaster BAC library, partial EcoRI in

PBACe3.6)"

BASE COUNT 54044 a 41124 c 41321 g 54184 t

ORIGIN

Query Match 75.2%; Score 18.8; DB 4; Length 190673;

Best Local Similarity 90.9%; Pred. No. 3.1e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 gatctgctgttgagcgagc 23

||||||| |||||||

Db 74835 GATCTGCTCTTGAAGCGCAGC 74814

## RESULT 30

```

AC025411
LOCUS       191736 bp      DNA             22-MAR-2000
DEFINITION  Homo sapiens chromosome X clone RP11-306E15 map X, WORKING DRAFT
SEQUENCE, 18 unordered pieces.
AC025411
VERSION     AC025411.2   GI:7283744
KEYWORDS    HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 191736)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL     Homo sapiens chromosome X, clone RP11-306E15
            Unpublished
REFERENCE   2 (bases 1 to 191736)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,P., Chepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Dekrellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Melidim,J., Meneus,L., Miranova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
            Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Mar 22, 2000 this sequence version replaced gi:7210134.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WtBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence.submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L8131
            Center clone name: 306.E.15
            ----- Summary Statistics
            Sequencing vector: M13; M77615; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 181491 bases at least Q40
            Consensus quality: 186630 bases at least Q30
            Consensus quality: 188547 bases at least Q20
            Insert size: 192000; agarose-fp
            Insert size: 190036; sum-of-contigs
            Quality coverage: 4.6 in Q20 bases; agarose-fp
            Quality coverage: 4.6 in Q20 bases; sum-of-contigs
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will

```

```

* be preserved.
1 1537: contig of 1537 bp in length
* 1538 1637: gap of 100 bp
* 1638 3266: contig of 1629 bp in length
* 3267 3366: gap of 100 bp
* 3367 7315: contig of 3949 bp in length
* 7316 7415: gap of 100 bp
* 7416 12732: contig of 5317 bp in length
* 12733 12832: gap of 100 bp
* 12833 18750: contig of 5918 bp in length
* 18751 18850: gap of 100 bp
* 18851 24921: contig of 6071 bp in length
* 24922 25021: gap of 100 bp
* 25022 30115: contig of 5094 bp in length
* 30116 30215: gap of 100 bp
* 30216 38744: contig of 8529 bp in length
* 38745 38844: gap of 100 bp
* 38845 45042: contig of 6198 bp in length
* 45043 45142: gap of 100 bp
* 45143 53343: contig of 8201 bp in length
* 53344 53443: gap of 100 bp
* 53444 59314: contig of 5871 bp in length
* 59315 59414: gap of 100 bp
* 59415 66278: contig of 6864 bp in length
* 66279 66378: gap of 100 bp
* 66379 73615: contig of 7237 bp in length
* 73616 73715: gap of 100 bp
* 73716 83850: contig of 10135 bp in length
* 83851 83950: gap of 100 bp
* 83951 95698: contig of 11748 bp in length
* 95699 95798: gap of 100 bp
* 95799 115031: contig of 19233 bp in length
* 115032 115131: gap of 100 bp
* 115132 143998: contig of 28867 bp in length
* 143999 144098: gap of 100 bp
* 144099 191736: contig of 47638 bp in length.

FEATURES
             Location/Qualifiers
             1..191736
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="X"
             /map="X"
             /clone="RP11-306E15"
             /clone_lib="RPC1-11 Human Male BAC"
             1..1537
             /note="assembly_fragment"
             1638..3266
             /note="assembly_fragment"
             3367..7315
             /note="assembly_fragment"
             7416..12732
             /note="assembly_fragment"
             clone_end:77
             vector_side:right"
             12833..18750
             /note="assembly_fragment"
             18851..24921
             /note="assembly_fragment"
             25022..30115
             /note="assembly_fragment"
             30216..38744
             /note="assembly_fragment"
             38845..45042
             /note="assembly_fragment"
             45143..53343
             /note="assembly_fragment"
             53444..59314
             /note="assembly_fragment"
             59415..66278
             /note="assembly_fragment"
             66379..73615
             /note="assembly_fragment"
             73716..83850
             /note="assembly_fragment"

```

misc\_feature 83951..95698 /note="assembly\_fragment"  
 misc\_feature 95799..115031 /note="assembly\_fragment"  
 misc\_feature 115132..143998 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:left  
 misc\_feature 144099..191736 /note="assembly\_fragment"  
 BASE COUNT 59272 a 36008 c 35590 g 59163 t 1703 others  
 ORIGIN

Query Match 75.2% Score 18.8; DB 69; Length 191736;  
 Best Local Similarity 90.9%; Pred. No. 3.1e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 ggatctgctgttgagcgag 22  
 |||||  
 Db 96915 GGATCTGCTGTCTGAAGCACAG 96936  
 |||||  
 RESULT 31  
 AE003686/c  
 LOCUS  
 DEFINITION Drosophila melanogaster genomic scaffold 142000013386035 section 11  
 of 105, complete sequence.  
 ACCESSION AE003686 AE002708  
 VERSION AE003686.2 GI:10726424  
 KEYWORDS HTG.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 221888)  
 Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,  
 Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,  
 George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,  
 Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,  
 Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,  
 Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor  
 Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,  
 Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Beeson,K.Y.,  
 Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,  
 Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,  
 Botchan,M.R., Bouck,J.J., Brokstein,P., Brottier,P., Burtis,K.C.,  
 Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,  
 Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de  
 Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,  
 Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,  
 Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,  
 Fleischmann,W., Fosler,K., Gabrielian,A.E., Garg,N.S.,  
 Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,  
 Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,  
 Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,  
 Wei,M.H., Jbergman,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,  
 Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,  
 Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,  
 Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,  
 McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Murrill,C.,  
 Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,  
 Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,  
 Nusskern,D.R., Pacle,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,  
 Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,  
 Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,  
 Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,  
 Stapleton,M., Strong,R., Sun,E., Svirskaas,R., Tector,C., Turner,R.,  
 Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,  
 Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,  
 Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,

Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,  
 Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,  
 Gibbs,R.A., Myers,E.W., Rubin,G.M., and Venter,J.C.  
 The genome sequence of Drosophila melanogaster  
 Science 287 (5461), 2185-2195 (2000)  
 20196006  
 REFERENCE  
 2 (bases 1 to 221888)  
 Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,J.C.  
 Direct Submission  
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD, USA  
 COMMENT  
 On Oct 9, 2000 this sequence version replaced gi:7299308.  
 Location/Qualifiers  
 1..221888  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /chromosome="3R"  
 complement(join(232..838,1727..2384,3247..3489,3547..4303,  
 4652..4762,5777..5989,8720..8991))  
 /gene="best"  
 /product="Ct19608"  
 /db\_xref="FLYBASE:FBan0006264"  
 /db\_xref="FLYBASE:FBgn0040238"  
 complement(<232..>8991)  
 /gene="best"  
 /note="CG6264"  
 /map="85F13-85F15"  
 /db\_xref="FLYBASE:FBan0006264"  
 /db\_xref="FLYBASE:FBgn0040238"  
 complement(join(513..838,1727..2384,3247..3489,3547..4303,  
 4652..4762,5777..5847))  
 /gene="best"  
 /note="best gene product"  
 /codon\_start=1  
 /db\_xref="FLYBASE:FBan0006264"  
 /db\_xref="FLYBASE:FBgn0040238"  
 /protein\_id="AAF54503.1"  
 /db\_xref="GI:7299309"  
 /translation="MTITYGEVATCRGFCGLKLLLRWRSIYKLVLDLAEFTIY  
 YAINMYRGLNPAQKTEFAIVQYCDSTRELIPSLFVLGFTVSVIMTRWQVTSIP  
 WPDPIAFVSSNVHGDGRMMRTIMRYVLCITMVLANSVPRKRPFGUNNLVE  
 AGLLNDENKTIETMNAKAPRPSKHLPIVMAASITTRAKEGRIDDPVAVKTIIDEL  
 NKFRGCGLLISYDTISVPLVTQVYTLAVSYVTCGQOQWTDGKVVGVNTLYLNKV  
 DLFYFVTTQDFEYFMGLKVAESLNPFGEDDDDFEVMVMVDNRNLOVSLYLVDEMHH  
 DPELLKQDWDFVFNELPYTIAERFRENHPEPTAKIEVPKNAFSTSSVRID  
 EMADASGHFSAGNKMRLDSSVLVSGLTSRVNTVASALKRFLSRDSDRPSGAT  
 PSQDPYKFPASASSASLSGVVGSATSGAKPAGSURITQQVIEVEQATITSMRAN  
 DDPNVMDIFAQTSAGTSGPLQPPAHSEVDIPSRPSPYNRAQSYEPNLFPPGG  
 VDALLSTAPAGGSLLLNAATAPSPVSGSSKSLYDPQKASRETVESMDLRSTD  
 LIGDAVQPEDEGDDFDKLKAREKEKLMROOKNLARTISTAPGMEATAVPMVMPV  
 NVAVQAQLOPVASSADLLAGGQDFSNSTMKSEDAINGS"  
 complement(join(<10693..10854,11008..11150,11212..11403,  
 11465..11562,11780..11858,11924..12033,13138..13244,  
 13478..>13489))  
 /gene="CG12814"  
 /product="Ct31942"  
 /db\_xref="FLYBASE:FBan0012814"  
 /db\_xref="FLYBASE:FBgn0037796"  
 /evidence=not\_experimental  
 complement(<10693..>13489)  
 /gene="CG12814"  
 /map="85F14-85F15"  
 /db\_xref="FLYBASE:FBan0012814"  
 /db\_xref="FLYBASE:FBgn0037796"  
 /evidence=not\_experimental  
 complement(join(10693..10854,11008..11150,11212..11403,  
 11465..11562,11780..11858,11924..12033,13138..13244,  
 13478..13489))  
 /gene="CG12814"  
 /note="CG12814 gene product"  
 /codon\_start=1  
 /db\_xref="FLYBASE:FBan0012814"  
 /db\_xref="FLYBASE:FBgn0037796"

```

/evidence-not_experimental
/protein_id="AAF54504.1"
/db_xref="GI:7299310"
/translation="MQLRTEERSIQLAVRVHKTLELADDKFYVITCGKSGYARDNNA
HVLKFLNDRHRETYVYGHYKIRAEFSKFNDFYGLRVGNCFAPDKKNTQKLTDD
GCPYDSKITSIFPTADGAAAEVLSMKFPEGSEVHLOCDVTCQYGRCEVDDCND
VALAGFGTNGPKRFGPNERGSLACTTVFVLDPAEALRSLNCEDGIRPSLLWLT
ITLGVFLIMLIMNIFLCTANSCSCANTEIIEKEPSIIEEDPYRSHWGSQYGRSVSL
HGRDAHKGTSGGSTRHSNRSRDY"
complement(<17074..>17286)
/gene="CG12815"
/product="CT31943"
/db_xref="FLYBASE:FBan0012815"
/db_xref="FLYBASE:FBgn0040542"
/evidence-not_experimental
complement(<17074..>17286)
/gene="CG12815"
/map="85F14-85F15"
/db_xref="FLYBASE:FBan0012815"
/db_xref="FLYBASE:FBgn0040542"
/evidence-not_experimental
complement(<17074..>17286)
/gene="CG12815"
/note="CG12815 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0012815"
/db_xref="FLYBASE:FBgn0040542"
/evidence-not_experimental
complement(<17074..>17286)
/protein_id="AAF54505.1"
/db_xref="GI:7299311"
/translation="MNKVRKMSGGYTGAVHVRDYRTPGCMAGDGGSDQVAFSLNLWAK
QAGSDCGILVSNVSGNSVTSALL"
join(<34263..34473,35044..35207,35267..35338,42493..42636,
42700..42822,51932..52066,53436..53677,54020..54116,
55551..55718,55778..56050,56213..57179,58267..59336)
/gene="Syn"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/product="CT13210"
/db_xref="FLYBASE:FBan0003985"
/db_xref="FLYBASE:FBgn0004575"
<34263..>59336
/gene="Syn"
/note="CG3985"
/map="85F15-85F16"
/db_xref="FLYBASE:FBan0003985"
/db_xref="FLYBASE:FBgn0004575"
join(34263..34473,35044..35207,35267..35338,42493..42636,
42700..42822,51932..52066,53436..53677,54020..54116,
55551..55718,55778..56050,56213..57179,58267..59631)
/gene="Syn"
/note="Syn gene product; Nucleotide sequence of the Celera
sequence differs from the published sequence for this
transcript"
/codon_start=1
/db_xref="FLYBASE:FBan0003985"
/db_xref="FLYBASE:FBgn0004575"
/protein_id="AAF54506.1"
/db_xref="GI:7299312"
/translation="MPPPPAPGAPGAAPFELSISFGAGTKPTATAAPAPRGRVSAPTSP
AKRESLLQRVQSLTGAAARDQGSILGAQVSATQAPAFSKDKYFTLLVLDQNTDW
SKYFRGLRQVDFDIREQAEFRDITVSSADTGPVVTMAAYRSRTPVARSFRPDPVL
IRQPPRDSGSDYRSTILGLKYGVPVINSLSIYQFDKPVWFHLLQLOLRRLGRDF
PLIOTFPPNPRDLISLRGMWRTHVLLRSCTFISYMLHLOHVAICTPDPVTPHY
RSTPRTEFLVSLHLOQFQKFPVSLKAGCHGCVATARENQSLQDAAGLVSGAGND
SHCYCTIEPYIDAKFSVHLQKIGNNYKAFMRKSTGNWNTNQSAMLEQIITETKYS
WVDEISELFGMEVCSVLVAVKAGREYIISACDSTFALIGDTEEDRQIADLVSGR
MQNVPRSMQATGPKLGRSSVSRAESPSTDEGVAPTPPLPAGRPAPMGPPPIPE
RTSPVAGTGRLSRSISSEVPSSSPSTVGGEIANKKRGTASSETSGSGPGSV
PSSAGPGSGFSSFLGKQFVAGKGEVISTOPTORPSEPPAIPATASSAVRPSVS
SVSDSRNTDITLTERAGAQPTVNTYEQQERVNPFKPEKPSKSGASIIHPSSSSISSS
SISRLNRNCAIQSPPPAGPPPTNTVAVGSNANSSSGYRNSFSSSLSKDKITSY
GNYGSTSVETITRMDTNTNIGATEAGEAGSVTALTNISNSDGIIVATPTGTITTS

```

```

VTNDRWSAIGMRASVYSAPAAVTIVLPDGTSGYDSNSIASQSGELNNPDLPSYTR
PSYSSESNAKSHSDLDIVFGDSKTTTPASYGNGKYTRAAGSISDADMTFGPPPSNYKT
DRFGASKSMSTSGVGSGSGGLGKYIYDSIQNAAFSDSFGSGMSSTGSHTKR
WSASKEDDELDK"
complement(join(42382..42513,42653..43196,43255..43473,
43581..43666,43738..43870,51247..51476))
/gene="timp"

Query Match      75.2%; Score 18.8; DB 5; Length 221888;
Best Local Similarity 90.9%; Pred. No. 3.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcgacg 23
||||| ||||||| ||||||| |||
Db 178558 GATCTGCTCTTGAAGCGACG 178537

RESULT 32
AB014726 1610 bp RNA VRL 12-MAR-1999
LOCUS Chuzan virus gene for outer capsid protein VP5, complete cds.
DEFINITION AB014726
ACCESSION AB014726
VERSION AB014726.1 GI:3986142
KEYWORDS outer capsid protein VP5.
SOURCE Chuzan virus cDNA to genomic RNA.
ORGANISM Chuzan virus
VIRUSES: dsRNA viruses; Reoviridae; Orbivirus; Palyam virus group.
REFERENCE 1 (sites)
AUTHORS Yamakawa,M., Furuuchi,S. and Minobe,Y.
TITLE Molecular characterization of double-stranded RNA segments encoding
the major capsid proteins of a Palyam serogroup orbivirus that
caused an epizootic of congenital abnormalities in cattle
J. Gen. Virol. 80 (Pt 1), 205-208 (1999)
99131402
REFERENCE 2 (bases 1 to 1610)
AUTHORS Yamakawa,M.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-1998) to the DDBJ/EMBL/GenBank databases. Makoto
Yamakawa, National Institute of Animal Health, Department of Exotic
Diseases, Josuihoncho 6-20-1, Kodaira, Tokyo 187-0022, Japan
(E-mail:yamato@ed.affrc.go.jp, Tel:+81-42-321-1441,
Fax:+81-42-325-5122)
FEATURES
Location/Qualifiers
1..1610
/organism="Chuzan virus"
/serotype="Palyam"
/db_xref="taxon:77204"
/chromosome="segment 6"
21..1586
/codon_start=1
/product="outer capsid protein VP5"
/protein_id="BAA34934.1"
/db_xref="GI:3986143"
/translation="MGREVKFLKRAKTKTMNALTNTAKSIYTKVTGKTIQKVASSELG
SAAIDGIIQGAARSALGEGNIGESIKQAVILNVMAANAPDPDISPGEHATYKRLAEL
EKEQEDRFPDQNEKELIDEGEELRSRQFGKMSILESGSEMSEMIKSIKMEK
IELKHTNDLOQLIRGLTKESMRNOREELISYFNNMRVLODAILSVQEGELHEEAVQ
IELDMGAELVETAAEEVPLVGAAGNAATASVRAVEGALKLVEIKLQSLGVDLSHMTYK
SLQPKMSLLRRNADGVITQKDLLEVNDKLLAVRELGDVKHLLLENIVPKTEKMY
DKQPKVGHAAAMKINKEFHPKIHITFAPMDSYVFMFKCVAPHHKEGFFLGFDEL
DPVYFEDLKVGAHOLIEGATEIVGRSFROIYRDFYFANVYSGASETHKKRLQRSSA
HPVILGSDYVOISIDOLYAHANQLVTNEELQLHLVRLPHFQRTINAAILLHVEIWT
RPNELTNDQAQSD"
BASE COUNT 555 a 237 c 418 g 400 t
ORIGIN

Query Match      74.4%; Score 18.6; DB 58; Length 1610;
Best Local Similarity 84.0%; Pred. No. 3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggaatctgctgttgaagcgacg 25
||||| | ||||||| |||

```



the expression of which is down-regulated by beta-catenin  
Cancer Res. 60 (13), 3354-3358 (2000)  
20365392  
2 (bases 1 to 6572)  
Nakamura, Y.  
Direct Submission  
Submitted (20-JUL-1999) Yusuke Nakamura, Institute of Medical  
Science, The University of Tokyo, Laboratory of Molecular Medicine,  
Human Genome Center; 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639,  
Japan (E-mail: zoe@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5372,  
Fax: 81-3-5449-5433)

## FEATURES

Location/Qualifiers

Source

1..6572

/organism="Mus musculus"

/db\_xref="taxon:10090"

1257..2822

/gene="Dctnmb1"

1257..2822

/note="Dctnmb1"

/note="down-regulated by ctnnb1, a"

/codon\_start=1

/product="Dctnmb1"

/protein\_id="BAB39850.1"

/db\_xref="GI:13442786"

/translation="MFTSEIGVVEELSEKTLPELTPNYATNLKDKSLVSLYKY  
IOEQSELLEPVCHOLFEPYSGEQLLRTFLQELPELMWCYLAVSRDVSQGLCY  
ALLGYNLEIVKRGHSLFTIPSLKSPVYHPSSIGSMALTESALSHGLSKV  
VYSGPHQREMLTQANRFEVLFTLLCYNALTYMPSVLSQSLCIGSCVCGYPRQ  
HVRYGVSSRIPISSGPMVQMLTGVIYFAYNGEWLAQALDDIYRAQLELYPEPL  
LVANAIKASLPHGAKSKKEGTCIQVEITPTSSRIKNAVTSMISIRGHWKRGDTE  
LTQGEELMDIPEVDEGYFSAASSTQSGLSNHNCNKTSTVGNORRSGSKAGAK  
ERETAGESCROHFARKOTQRAQSENLELLSLRLTLTSSLOPKPSQGLAKTAATVF  
SKSPQVSGAPVPSPGTCVAGADANRFSACSLQELIIVSERTELAVKQCAQO  
OGPSSIVTLSEAE"

BASE COUNT 1760 a 1352 c 1548 g 1904 t 8 others

ORIGIN

Query Match 74.4%; Score 18.6; DB 94; Length 6572;  
Best Local Similarity 84.0%; Pred. No. 3.2e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gqatctgtctgttgtaagcgcagcag 25

||||| | ||||| |||||

Db 1491 GGAAGTCGACAGTGAAGCGCAGCAG 1467

RESULT 36

AE005138

LOCUS

AE005138

DEFINITION

Halobacterium sp. NRC-1

ACCESSION

AE005138.1

VERSION

GI:10582037

KEYWORDS

Halobacterium sp. NRC-1.

SOURCE

Halobacterium sp. NRC-1.

ORGANISM

Halobacterium sp. NRC-1.

REFERENCE

1 (bases 1 to 10529)

Ng, W.V., Kennedy, S.P., Mahairas, G.G., Berquist, B., Pan, M.,

Shukla, H.D., Lasky, S.R., Balliga, N., Thorsson, V., Sbrogna, J.,

Swartzell, S., Weir, D., Hall, J., Dahl, T.A., Welti, R., Goo, Y.A.,

Leithauser, B., Keller, K., Cruz, R., Danson, M.J., Hough, D.W.,

Maddocks, D.G., Jablonski, P.E., Krebs, M.P., Angevine, C.M., Dale, H.,

Isenbarger, T.A., Peck, R.F., Pohlschrod, M., Spudich, J.L.,

Jung, K.-H., Alam, M., Freitas, T., Hou, S., Daniels, C.J., Dennis, P.P.,

Omer, A.D., Ebhardt, H., Lowe, T.M., Liang, P., Riley, M., Hood, L. and

DasSarma, S.

From the cover: genome sequence of halobacterium species NRC-1

Proc. Natl. Acad. Sci. USA 97 (22), 12176-12181 (2000)

PUBMED

11016950

REFERENCE

2 (bases 1 to 10529)

Ng, W.V., Kennedy, S.P., Mahairas, G.G., Berquist, B., Pan, M.,

Shukla, H.D., Lasky, S.R., Balliga, N., Thorsson, V., Sbrogna, J.,  
Swartzell, S., Weir, D., Hall, J., Dahl, T.A., Welti, R., Goo, Y.A.,  
Leithauser, B., Keller, K., Cruz, R., Danson, M.J., Hough, D.W.,  
Maddocks, D.G., Jablonski, P.E., Krebs, M.P., Angevine, C.M., Dale, H.,  
Isenbarger, T.A., Peck, R.F., Pohlschrod, M., Spudich, J.L.,  
Jung, K.-H., Alam, M., Freitas, T., Hou, S., Daniels, C.J., Dennis, P.P.,  
Omer, A.D., Ebhardt, H., Lowe, T.M., Liang, P., Riley, M., Hood, L. and  
DasSarma, S.

TITLE

JOURNAL

FEATURES

Source

1..10529

/organism="Halobacterium sp. NRC-1"

/strain="NRC-1"

/db\_xref="taxon:64091"

196..960

/gene="VNG2656H"

196..960

/note="hypothetical protein"

/codon\_start=1

/transl\_table=11

/product="Vng2656h"

/protein\_id="AAG20687.1"

/db\_xref="GI:10582038"

/translation="MEIRVSGAGPAAPFSSARDVFEPSHDLDPVDVHIRSNPDERTW  
AGHYDDHVLNI SQQAATSVMARELALHEFAIMHRHCHPSHTLSMDEVLFALTGR  
EVEQRLTHCYOIAANHVKD IYADDITLSVGPDTDKLSRFLSELAALADQADQOAG  
HRLTAGADSPMTAVNAAFALALHEDALIPGDRHLYDLAHAGQDAPEIDVDGFRFRF  
ATLPPDPSDCRRGLVDTIRYADAQTITTPAAD"

/note="VNG2657G"

/complement(1035..1667)

/gene="tps7p"

/note="tps7p"

/codon\_start=1

/transl\_table=11

/product="30S ribosomal protein S7p"

/protein\_id="AAG20688.1"

/db\_xref="GI:10582039"

/translation="MSDEQPAEDETEAAFESEDTOEVAANAKLFGKWDVAEIHVEDP  
STRYLAVTPVAHTMGRHAQKQFKSEISIVERLANRLMKCANAGKQOALKIVRDA  
FDIVHRTDENP IQVLNSAVENAPRETIVRLKGGISVPOAVDPAPORRVQDQALKFL  
ADGHSASFKPTDAAALAGADYNNQVTAIGQKKEKERVAAAA"

/complement(1671..2099)

/gene="tps12p"

/note="tps12p"

/codon\_start=1

/transl\_table=11

/product="30S ribosomal protein S12p"

/protein\_id="AAG20689.1"

/db\_xref="GI:10582040"

/translation="MTNCKYARLKKDQORRWSDEYARERGLCKSDPLEGAPQ  
GRGIVLEKVGIEAKQPNRAIKRVQVLIQKNGQVAFPCPGDGAISFIDEHDEVTIAG  
IGGAKGRAMDLSGVNYKVKVGVSLIELVRGNAKEPVR"

/complement(2263..2682)

/gene="nusa"

/note="nusa"

/complement(2263..2682)

/gene="nusa"

/note="nusa"

/codon\_start=1

/transl\_table=11

/product="transcription termination-antitermination  
factor"

/protein\_id="AAG20690.1"

/db\_xref="GI:10582041"

/translation="MTIRLSDEARQLIAVFETETATATVDCLPDDDRVVVVVTAGEMG



```

/clone_lib="Lambda EMBL3"
1351..1356
/feature="promoter 1"
/translation="MSAGQAPKEIGSFGLMDPEEVRDMSATKVITADTYDDGGFPI
DMGLMDPRLGVIDPGECKTCGQSGCNGHFGHIELAAPVIHVGFSLIRLLRGTG
REKASILLTEEEKDEYENLDRTSLRQDVSDVMTAAIRPARKKDHCHPCGGEVQYDVK
HEKPTTYVEQVQVLAADYSERIAASMOPEDEDDAGVQELAEQTDIDISRIEILS
GEFRPREDEAREIETALGADLTEDMKNLMPSDIRWFDIEDIGEDLEALGNSDRSP
EMMLTLVPPVPTNPSTLNGQSRSEDDTHKLVDIIRINOFNREAGAPQLTI
EDLWELQVHTVTFMDNEISGTPPARHSRGLTKLSQRLKGEKGRFSGLSGRVNF
SARTVISPDTLSLNEGVPRVATEMTQVNVNEQLNFRARYVRNGEGHPGANYV
TRPDGRVRVTEKVCBELAEPVGEVQVHILIDGDIILFNROPSLHRMSIMAEVVV
MPYFRLNVTVCPPYNADEFDGMNHALQNEEARAEARVLMRVOEQLISPRFGENI
IGAIOHISCTYLLTNDNPRENETQASDLRLRUIDELPAAGTDEDDGQYWTGHQIF
SELLPDDLSLEFTGTGTDVWIEDGQELLEGTIADDEVGEYGEIVDTITIKVHGTRAR
IFINVASLAMSRIHMFSGISIGIDDETSTEARERIDEAIOAIDRVQELIETENG
LESIGRTVDELTLEMKIMOTLGKARDSAGDAENFEDENPAVMANSARGSMNLIT
MGREGIVDTAVRTSKSGVIOQLRLINALSELTQYDGTVRDTSITVQEPFGEDGTSP
VOVSSNEEVIDVEHVADRILLNSEPDSDTQKAEFLVEVEPPTNLSEHGAWEVESDD"
5050..7962
/feature="RNA polymerase subunit operon"
/product="RNA polymerase subunit A"
7958..9148
/feature="RNA polymerase subunit operon"
/product="RNA polymerase subunit C"
7958..9151
/feature="RNA polymerase subunit operon"
/codon_start=1
/transl_table=11
/product="RNA polymerase subunit C"
/protein_id="CAA40427.1"
/db_xref="GI:43543"
/db_xref="SWISS-PROT:P15354"
/translation="MTFNTVDTGVHDVIEDTELPRRLKDEVLTAEERCVTKSOANEI
ATAVEAQYLDTRVLDLPVGTYSAGSICEPGCTQMTMNTFHYAGVEMDVTQGLPRLIE
LVDAKPTPTVMEVLTDEYEAERAEHVEVVKIETAKILAGDIDSTNVAWVRID
LNEDTQERWPRVDSITQAGVAETIEGNLGVTVTQDGTILEFGPSPSYRELLQLV
EOLRDVFKGIEEVRIRRETEERGEFVLYTEGSAFKALKIEGSDATSCNNI
HEVHTLGLAEAREALINFTMTLEEQGLDDVNIIRMLVADIMTNDGTIESIGHGI
SGNKSVLARAAFEVTVNHLDAATHGESDDLDGVNIENVIVGKPVRLGTGDVDRMGA
TQTSQ"
9151..9567
/feature="ORF 139"
9151..9570
/feature="ORF 139"
9151..9570
/feature="ORF 139"
/codon_start=1
/transl_table=11
/protein_id="CAA40428.1"
/db_xref="GI:43544"
/db_xref="SWISS-PROT:P15739"
/translation="MTIRLSDEAROLIAVFETETATVDCILPDDRRVVVVYTAGEMG
AAIGDGSRVDALEATLGRSVVLVEDAPTAEGFVANALSPAAYNVNTVSENDTTVAYA
EVAHEDKGAICADGTNIETAKELARHEDIDDIQLT"
9705..9710
/feature="promoter 2"
9734..10159
/feature="RNA polymerase subunit operon"
/product="ribosomal protein Hhs12"
9734..10162
/feature="RNA polymerase subunit operon"
9734..10162
/feature="RNA polymerase subunit operon"
/codon_start=1
/transl_table=11
/product="ribosomal protein Hhs12"
/protein_id="CAA40429.1"
/db_xref="GI:43545"
/db_xref="SWISS-PROT:P15756"
/translation="MTNGYAAKRLKDKRQORRWSDSYARRERCLGKKSPLLEGAPQ
GHGIVLEKVGIEAKQPNIAKRCVKVLIKNKGQVTAFCPGDGAIISFIDEHDEVTIAG
IGGAKGRAMGDLSGVNYKVEKVGVSILIELVRGNAEKPR"
1351..1356
/feature="promoter 1"
/translation="MSAGQAPKEIGSFGLMDPEEVRDMSATKVITADTYDDGGFPI
DMGLMDPRLGVIDPGECKTCGQSGCNGHFGHIELAAPVIHVGFSLIRLLRGTG
REKASILLTEEEKDEYENLDRTSLRQDVSDVMTAAIRPARKKDHCHPCGGEVQYDVK
HEKPTTYVEQVQVLAADYSERIAASMOPEDEDDAGVQELAEQTDIDISRIEILS
GEFRPREDEAREIETALGADLTEDMKNLMPSDIRWFDIEDIGEDLEALGNSDRSP
EMMLTLVPPVPTNPSTLNGQSRSEDDTHKLVDIIRINOFNREAGAPQLTI
EDLWELQVHTVTFMDNEISGTPPARHSRGLTKLSQRLKGEKGRFSGLSGRVNF
SARTVISPDTLSLNEGVPRVATEMTQVNVNEQLNFRARYVRNGEGHPGANYV
TRPDGRVRVTEKVCBELAEPVGEVQVHILIDGDIILFNROPSLHRMSIMAEVVV
MPYFRLNVTVCPPYNADEFDGMNHALQNEEARAEARVLMRVOEQLISPRFGENI
IGAIOHISCTYLLTNDNPRENETQASDLRLRUIDELPAAGTDEDDGQYWTGHQIF
SELLPDDLSLEFTGTGTDVWIEDGQELLEGTIADDEVGEYGEIVDTITIKVHGTRAR
IFINVASLAMSRIHMFSGISIGIDDETSTEARERIDEAIOAIDRVQELIETENG
LESIGRTVDELTLEMKIMOTLGKARDSAGDAENFEDENPAVMANSARGSMNLIT
MGREGIVDTAVRTSKSGVIOQLRLINALSELTQYDGTVRDTSITVQEPFGEDGTSP
VOVSSNEEVIDVEHVADRILLNSEPDSDTQKAEFLVEVEPPTNLSEHGAWEVESDD"
5050..7962
/feature="RNA polymerase subunit operon"
/product="RNA polymerase subunit A"
7958..9148
/feature="RNA polymerase subunit operon"
/product="RNA polymerase subunit C"
7958..9151
/feature="RNA polymerase subunit operon"
/codon_start=1
/transl_table=11
/product="RNA polymerase subunit C"
/protein_id="CAA40427.1"
/db_xref="GI:43543"
/db_xref="SWISS-PROT:P15354"
/translation="MTFNTVDTGVHDVIEDTELPRRLKDEVLTAEERCVTKSOANEI
ATAVEAQYLDTRVLDLPVGTYSAGSICEPGCTQMTMNTFHYAGVEMDVTQGLPRLIE
LVDAKPTPTVMEVLTDEYEAERAEHVEVVKIETAKILAGDIDSTNVAWVRID
LNEDTQERWPRVDSITQAGVAETIEGNLGVTVTQDGTILEFGPSPSYRELLQLV
EOLRDVFKGIEEVRIRRETEERGEFVLYTEGSAFKALKIEGSDATSCNNI
HEVHTLGLAEAREALINFTMTLEEQGLDDVNIIRMLVADIMTNDGTIESIGHGI
SGNKSVLARAAFEVTVNHLDAATHGESDDLDGVNIENVIVGKPVRLGTGDVDRMGA
TQTSQ"
9151..9567
/feature="ORF 139"
9151..9570
/feature="ORF 139"
9151..9570
/feature="ORF 139"
/codon_start=1
/transl_table=11
/protein_id="CAA40428.1"
/db_xref="GI:43544"
/db_xref="SWISS-PROT:P15739"
/translation="MTIRLSDEAROLIAVFETETATVDCILPDDRRVVVVYTAGEMG
AAIGDGSRVDALEATLGRSVVLVEDAPTAEGFVANALSPAAYNVNTVSENDTTVAYA
EVAHEDKGAICADGTNIETAKELARHEDIDDIQLT"
9705..9710
/feature="promoter 2"
9734..10159
/feature="RNA polymerase subunit operon"
/product="ribosomal protein Hhs12"
9734..10162
/feature="RNA polymerase subunit operon"
9734..10162
/feature="RNA polymerase subunit operon"
/codon_start=1
/transl_table=11
/product="ribosomal protein Hhs12"
/protein_id="CAA40429.1"
/db_xref="GI:43545"
/db_xref="SWISS-PROT:P15756"
/translation="MTNGYAAKRLKDKRQORRWSDSYARRERCLGKKSPLLEGAPQ
GHGIVLEKVGIEAKQPNIAKRCVKVLIKNKGQVTAFCPGDGAIISFIDEHDEVTIAG
IGGAKGRAMGDLSGVNYKVEKVGVSILIELVRGNAEKPR"

```





Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L6593  
 Center clone name: 2017\_B\_4  
 -----

\* NOTE: This record contains 83 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1 737: contig of 737 bp in length  
 738 837: gap of 100 bp  
 838 1603: contig of 766 bp in length  
 1604 1703: gap of 100 bp  
 1704 2455: contig of 752 bp in length  
 2456 2555: gap of 100 bp  
 2556 3321: contig of 766 bp in length  
 3322 3421: gap of 100 bp  
 3422 4165: contig of 744 bp in length  
 4166 4265: gap of 100 bp  
 4266 5016: contig of 751 bp in length  
 5017 5116: gap of 100 bp  
 5117 5868: contig of 752 bp in length  
 5869 5968: gap of 100 bp  
 5969 6704: contig of 736 bp in length  
 6705 6804: gap of 100 bp  
 6805 7561: contig of 757 bp in length  
 7562 7661: gap of 100 bp  
 7662 8408: contig of 747 bp in length  
 8409 8508: gap of 100 bp  
 8509 9255: contig of 747 bp in length  
 9256 9355: gap of 100 bp  
 9356 10116: contig of 761 bp in length  
 10117 10216: gap of 100 bp  
 10217 10967: contig of 751 bp in length  
 10968 11067: gap of 100 bp  
 11068 11819: contig of 752 bp in length  
 11820 11919: gap of 100 bp  
 11920 12677: contig of 758 bp in length  
 12678 12777: gap of 100 bp  
 12778 13519: contig of 742 bp in length  
 13520 13619: gap of 100 bp  
 13620 14352: contig of 733 bp in length  
 14353 14452: gap of 100 bp  
 14453 15206: contig of 754 bp in length  
 15207 15306: gap of 100 bp  
 15307 16056: contig of 750 bp in length  
 16057 16156: gap of 100 bp  
 16157 16921: contig of 765 bp in length  
 16922 17021: gap of 100 bp  
 17022 17766: contig of 745 bp in length  
 17767 17866: gap of 100 bp  
 17867 18601: contig of 735 bp in length  
 18602 18701: gap of 100 bp  
 18702 19458: contig of 757 bp in length  
 19459 19558: gap of 100 bp  
 19559 20312: contig of 754 bp in length  
 20313 20412: gap of 100 bp  
 20413 21169: contig of 757 bp in length  
 21170 21269: gap of 100 bp  
 21270 22012: contig of 743 bp in length  
 22013 22112: gap of 100 bp  
 22113 22857: contig of 745 bp in length  
 22858 22957: gap of 100 bp  
 22958 23697: contig of 740 bp in length  
 23698 23797: gap of 100 bp  
 23798 , 24544: contig of 747 bp in length

24545 24644: gap of 100 bp  
 24645 25399: contig of 755 bp in length  
 25400 25499: gap of 100 bp  
 25500 26246: contig of 747 bp in length  
 26247 26346: gap of 100 bp  
 26347 27105: contig of 759 bp in length  
 27106 27205: gap of 100 bp  
 27206 27943: contig of 738 bp in length  
 27944 28043: gap of 100 bp  
 28044 28793: contig of 750 bp in length  
 28794 28893: gap of 100 bp  
 28894 29648: contig of 755 bp in length  
 29649 29748: gap of 100 bp  
 29749 30510: contig of 762 bp in length  
 30511 30610: gap of 100 bp  
 30611 31355: contig of 745 bp in length  
 31356 31455: gap of 100 bp  
 31456 32187: contig of 732 bp in length  
 32188 32287: gap of 100 bp  
 32288 33032: contig of 745 bp in length  
 33033 33132: gap of 100 bp  
 33133 33884: contig of 752 bp in length  
 33885 33984: gap of 100 bp  
 33985 34729: contig of 745 bp in length  
 34730 34829: gap of 100 bp  
 34830 35573: contig of 744 bp in length  
 35574 35673: gap of 100 bp  
 35674 36427: contig of 754 bp in length  
 36428 36527: gap of 100 bp  
 36528 37281: contig of 754 bp in length  
 37282 37381: gap of 100 bp  
 37382 38149: contig of 768 bp in length  
 38150 38249: gap of 100 bp  
 38250 39005: contig of 756 bp in length  
 39006 39105: gap of 100 bp  
 39106 39852: contig of 747 bp in length  
 39853 39952: gap of 100 bp  
 39953 40698: contig of 746 bp in length  
 40699 40798: gap of 100 bp  
 40799 41554: contig of 756 bp in length  
 41555 41654: gap of 100 bp  
 41655 42411: contig of 757 bp in length  
 42412 42511: gap of 100 bp  
 42512 43252: contig of 741 bp in length  
 43253 43352: gap of 100 bp  
 43353 44112: contig of 760 bp in length  
 44113 44212: gap of 100 bp  
 44213 44964: contig of 752 bp in length  
 44965 45064: gap of 100 bp  
 45065 45810: contig of 746 bp in length  
 45811 45910: gap of 100 bp  
 45911 46672: contig of 762 bp in length  
 46673 46772: gap of 100 bp  
 46773 47531: contig of 759 bp in length  
 47532 47631: gap of 100 bp  
 47632 48375: contig of 744 bp in length  
 48376 48475: gap of 100 bp  
 48476 49222: contig of 747 bp in length  
 49223 49322: gap of 100 bp  
 49323 50066: contig of 744 bp in length  
 50067 50166: gap of 100 bp  
 50167 50911: contig of 745 bp in length  
 50912 51011: gap of 100 bp  
 51012 51766: contig of 755 bp in length  
 51767 51866: gap of 100 bp  
 51867 52624: contig of 758 bp in length  
 52625 52724: gap of 100 bp  
 52725 53476: contig of 752 bp in length  
 53477 53576: gap of 100 bp  
 53577 54314: contig of 738 bp in length  
 54315 54414: gap of 100 bp  
 54415 55155: contig of 741 bp in length  
 55156 55255: gap of 100 bp

\* 55256 56023: contig of 768 bp in length  
 \* 56024 56123: gap of 100 bp  
 \* 56124 56868: contig of 745 bp in length  
 \* 56869 56968: gap of 100 bp  
 \* 56969 57693: contig of 725 bp in length  
 \* 57694 57793: gap of 100 bp  
 \* 57794 58544: contig of 751 bp in length  
 \* 58545 58644: gap of 100 bp

Query Match 74.4%; Score 18.6; DB 73; Length 70326;  
 Best Local Similarity 84.0%; Pred. No. 3.7e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggatctgtgttgagcgagcag 25  
 |||||  
 DB 58138 GGATCTGCTGTTGGAAAGCTGCAG 58114

## RESULT 40

AC022946 150516 bp DNA HTG 28-MAR-2000  
 LOCUS Homo sapiens chromosome 21 clone RP11-70H18 map 21, WORKING DRAFT  
 DEFINITION SEQUENCE, 15 unordered pieces.  
 AC022946  
 VERSION AC022946.2 GI:7331470  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 150516)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 21, clone RP11-70H18

Unpublished

2 (bases 1 to 150516)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,  
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,  
 Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,  
 DeArrellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,  
 Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,  
 Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Landers,T., Lehoczy,J., Levine,R., Liu,C., Locke,K.,  
 MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
 McPheeters,R., Meldrim,J., Menus,L., Morrow,J., Naylor,J.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,  
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
 Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
 Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
 Zimmer,A. and Zody,M.

Direct Submission

Submitted (07-FEB-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 28, 2000 this sequence version replaced gi:6921811.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L6396

Center clone name: 70\_H.18

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 142388 bases at least Q40

Consensus quality: 146212 bases at least Q30

Consensus quality: 148009 bases at least Q20  
 Insert size: 148000; agarose-fp  
 Insert size: 149116; sum-of-contigs  
 Quality coverage: 4.8 in Q20 bases; agarose-fp  
 Quality coverage: 4.8 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 15 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 187: contig of 187 bp in length  
 188 287: gap of 100 bp  
 288 1328: contig of 1041 bp in length  
 1329 1428: gap of 100 bp  
 1429 3227: contig of 1799 bp in length  
 3228 3327: gap of 100 bp  
 3328 5694: contig of 2367 bp in length  
 5695 5794: gap of 100 bp  
 5795 8295: contig of 2501 bp in length  
 8296 8395: gap of 100 bp  
 8396 12308: contig of 3913 bp in length  
 12309 12408: gap of 100 bp  
 12409 17784: contig of 5376 bp in length  
 17785 17884: gap of 100 bp  
 17885 24561: contig of 6677 bp in length  
 24562 24661: gap of 100 bp  
 24662 31935: contig of 7274 bp in length  
 31936 32035: gap of 100 bp  
 32036 41289: contig of 9254 bp in length  
 41290 41389: gap of 100 bp  
 41390 48680: contig of 7291 bp in length  
 48681 48780: gap of 100 bp  
 48781 60172: contig of 11392 bp in length  
 60173 60272: gap of 100 bp  
 60273 74915: contig of 14643 bp in length  
 74916 75015: gap of 100 bp  
 75016 107000: contig of 31985 bp in length  
 107001 107100: gap of 100 bp  
 107101 150516: contig of 43416 bp in length.

## FEATURES

source

1..150516  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="21"  
 /map="21"  
 /clone="RP11-70H18"  
 /clone\_lib="RPC1-11 Human Male BAC"

misc\_feature

1..187  
 /note="assembly\_fragment  
 clone\_end:T7  
 vector\_side:right"

misc\_feature

288..1328  
 /note="assembly\_fragment"  
 1429..3227  
 /note="assembly\_fragment"

misc\_feature

3328..5694  
 /note="assembly\_fragment"  
 5795..8295  
 /note="assembly\_fragment"

misc\_feature

8396..12308  
 /note="assembly\_fragment"  
 12409..17784  
 /note="assembly\_fragment"

misc\_feature

17885..24561  
 /note="assembly\_fragment"  
 24662..31935  
 /note="assembly\_fragment"

misc\_feature

32036..41289  
 /note="assembly\_fragment"

```

misc_feature 41390..48680
/note="assembly_fragment
clone_end:SP6
vector_side:right"
misc_feature 48781..60172
/note="assembly_fragment"
misc_feature 60273..74915
/note="assembly_fragment"
misc_feature 75016..107000
/note="assembly_fragment"
misc_feature 107101..150516
/note="assembly_fragment"
BASE COUNT 49855 a 26387 c 27406 g 45468 t 1400 others
ORIGIN

Query Match 74.4%; Score 18.6; DB 67; Length 150516;
Best Local Similarity 84.0%; Pred No. 3.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggatctgctgttgaagcgcagcag 25
|||||
Db 30539 GGATCTGCTGTTGGAAAGCTGCAG 30563

RESULT 41
AF240629/C AF240629 151696 bp DNA PRI 19-MAR-2000
LOCUS Homo sapiens chromosome 21 map 21q21 clone B47C12, complete
DEFINITION
sequence.
ACCESSION AF240629
VERSION AF240629.1 GI:7263185
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 151696)
AUTHORS Blechschmidt,K., Riesselmann,L., Wehrmeyer,S., Baumgart,C.,
Menzel,U., Dette,M., Jahn,N., Schilhabel,M., Yaspo,M.-L. and
Rosenthal,A.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2000) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
source
1..151696
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21"
/clone="B47C12"
repeat_region 1..1947
/rpt_family="L1M1_5end"
repeat_region 1681..3013
/rpt_family="L1M2_orf2"
repeat_region 3009..3284
/rpt_family="L1M2_orf2"
repeat_region 3135..3756
/rpt_family="L1M3"
repeat_region 3754..3867
/rpt_family="L1M2_orf2"
repeat_region complement(3876..3939)
/rpt_family="L1M1"
repeat_region 3943..4342
/rpt_family="L1M1"
repeat_region 4358..6349
/rpt_family="Tigger1"
exon 4649..4773
/note="GRAIL"
/evidence=not_experimental
repeat_region complement(5163..5470)
/rpt_family="AluJo"
repeat_region complement(5871..6145)

```

```

exon /rpt_family="AluSc"
complement(6282..6346)
/note="MZF"
/evidence=not_experimental
complement(6630..6689)
/note="(TA)n"
/rpt_family="Simple_repeat"
6912..7214
/rpt_family="THE1C"
7241..8157
/rpt_family="THE1-INTERNAL"
7605..7838
/note="GENSCAN"
7822..7838
/note="XPOUND"
/evidence=not_experimental
8158..8538
/rpt_family="THE1C"
8703..8891
/rpt_family="L1MD2"
8737..8891
/rpt_family="L1MC2"
9211..9239
/note="AT-rich"
/rpt_family="Low_complexity"
10062..10156
/note="(CATA)n"
/rpt_family="Simple_repeat"
10427..10531
/note="MZF"
/evidence=not_experimental
10448..10943
/rpt_family="LTR22"
11180..11252
/note="GRAIL"
/evidence=not_experimental
11180..11252
/note="MZF"
/evidence=not_experimental
11180..11252
/note="GENSCAN"
12184..12573
/rpt_family="LTR40a"
complement(12677..12773)
/rpt_family="L2a"
12796..12875
/note="AT-rich"
/rpt_family="Low_complexity"
complement(13203..13221)
/note="XPOUND"
/evidence=not_experimental
complement(13313..13679)
/rpt_family="L2"
13549..13649
/note="GRAIL"
/evidence=not_experimental
complement(16076..16142)
/rpt_family="MSTA"
complement(16143..17744)
/rpt_family="MST-INTERNAL"
complement(17745..18149)
/rpt_family="MSTA"
complement(18237..18613)
/rpt_family="MSTA"
complement(20615..20739)
/note="GRAIL"
/evidence=not_experimental
20673..20725
/rpt_family="U2"
complement(20934..21068)
/note="MZF"
/evidence=not_experimental
20954..21248

```

```

/rpt_family="AluJb"
21393..21437
/note="GAIL"
/evidence=not_experimental
repeat_region
21777..21812
/rpt_family="L1PA3"
repeat_region
21815..21840
/note="AT-rich"
/rpt_family="Low_complexity"
21995..22142
/rpt_family="L1MD2"
complement(22044..22097)
/exon
/note="GAIL"
/evidence=not_experimental
complement(22048..22097)
/exon
/note="MZF"
/evidence=not_experimental
22287..22436
/note="MZF"
/exon
/evidence=not_experimental
complement(22312..22414)
repeat_region
/rpt_family="FLAM_A"
22446..22486
/exon
/note="CA)n"
/evidence=not_experimental
complement(23834..23878)
repeat_region
/rpt_family="Simple_repeat"
complement(23981..24687)
repeat_region
/rpt_family="LTR8"
25470..25541
/exon
/rpt_family="MER5B"
complement(25769..26077)
repeat_region
/rpt_family="AluSg"
complement(25979..26084)
/exon
/note="MZF"
/evidence=not_experimental
complement(26439..26499)
repeat_region
/note="(TA)n"
/rpt_family="Simple_repeat"
26526..26623
/exon
/note="MZF"
/evidence=not_experimental
complement(26982..27088)
/exon
/note="MZF"
/evidence=not_experimental
27646..27683
/exon
/note="MZF"
/evidence=not_experimental
27926..28316
repeat_region
/rpt_family="MLT1A2"
complement(28698..28721)
repeat_region
/note="AT-rich"
/rpt_family="Low_complexity"
complement(28989..29086)
repeat_region
/note="(TA)n"
/rpt_family="Simple_repeat"
complement(29452..29895)
repeat_region
/rpt_family="MER31B"
29515..29615
/exon
/note="GAIL"
/evidence=not_experimental
29970..30011
repeat_region
/note="AT-rich"
/rpt_family="Low_complexity"
30034..30055
repeat_region
/note="AT-rich"
/rpt_family="Low_complexity"
30214..30544
repeat_region
/note="AT-rich"
/rpt_family="Low_complexity"
30223..30508
repeat_region
/rpt_family="AluSg"
30519..30539
repeat_region
/note="(TAAA)n"
/rpt_family="Simple_repeat"
30585..30685
repeat_region
/note="(TA)n"
/rpt_family="Simple_repeat"
complement(31020..31380)
repeat_region
/rpt_family="MLT1A1"

Query Match 74.4%; Score 18.6; DB 89; Length 151696;
Best Local Similarity 84.0%; Pred. No. 3.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggatctctgtttgaagcgacgacg 25
|||||
Db 125637 GGATCTGCTGTTGGAAAGCTGCAG 125613
|||||

RESULT 42
AL138781 154923 bp DNA HTG 23-JAN-2001
LOCUS Homo sapiens chromosome 9 clone RP11-83N9 map p34.1-35.1, ***
DEFINITION SEQUENCING IN PROGRESS ***, 8 unordered pieces.
ACCESSION AL138781
VERSION AL138781.5 GI:9796232
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 154923)
AUTHORS Plumb, B.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 11, 2000 this sequence version replaced gi:9212004.
COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: dj1079D1
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 153405 bases at least Q40
Consensus quality: 153765 bases at least Q30
Consensus quality: 153962 bases at least Q20
Insert size: 154223; sum-of-contigs
Insert size: 137931; 10.9% error; agarose-fp
Quality coverage: 7.20x in Q20 bases; sum-of-contigs Quality
coverage: 8.52x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 9663: contig of 9663 bp in length
* 9664 9763: gap of 100 bp
* 9764 23806: contig of 14043 bp in length
* 23807 23906: gap of 100 bp
* 23907 43080: contig of 19174 bp in length
* 43081 43180: gap of 100 bp
* 43181 91899: contig of 48719 bp in length
* 91900 91999: gap of 100 bp
* 92000 103980: contig of 11981 bp in length
```

```
* 103981 104080: gap of 100 bp
* 104081 118518: contig of 14438 bp in length
* 118519 118618: gap of 100 bp
* 118619 149181: contig of 30563 bp in length
* 149182 149281: gap of 100 bp
* 149282 154923: contig of 5642 bp in length.
```

## FEATURES

```
source
1. .154923
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="p34.1-35.1"
/clone="RP11-83N9"
/clone_lib="RP11-11.1"
1. .9663
/note="assembly fragment:00750
fragment_chain:1
clone_end:SP6
vector_side:left"
9764. .23806
/note="assembly fragment:01482
fragment_chain:1"
23907. .43080
/note="assembly fragment:01923
fragment_chain:1"
43181. .91899
/note="assembly fragment:01849
fragment_chain:1"
92000. .103980
/note="assembly fragment:00203
fragment_chain:1"
104081. .118518
/note="assembly fragment:01769
fragment_chain:1"
118619. .149181
/note="assembly fragment:01437.0"
149282. .154923
/note="assembly fragment:02700"
702 others
```

```
Query Match 74.4%; Score 18.6; DB 78; Length 154923;
Best Local Similarity 84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 gqatctgttgaagcgcagcag 25
||||| | | | | | | | | |
Db 111365 GGATCTGCTCTGAAGCCGCGCAG 111389
```

## RESULT 43

```
AC011010/C
LOCUS AC011010 170896 bp DNA HTG 16-MAR-2000
DEFINITION Homo sapiens clone RP11-6P16, WORKING DRAFT SEQUENCE, 21 unordered
pieces.
ACCESSION AC011010
VERSION AC011010.4 GI:7107881
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 170896)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collumore,A.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
```

Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Lehoczky,J., Liew,C., Locker,K., Macdonald,P., Marquis,N.,  
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Melchior,J.,  
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

## TITLE

## JOURNAL

## COMMENT

Submitted (29-SEP-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Feb 28, 2000 this sequence version replaced gi:6479051.  
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RW/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L2916

Center clone name: 6\_P16

----- Summary Statistics

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 114103 bases at least Q40

Consensus quality: 141555 bases at least Q30

Insert size: 154000; agarose-fp

Quality coverage: 168896; sum-of-contigs

Quality coverage: 3.6 in Q20 bases; agarose-fp

Quality coverage: 3.3 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 21 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

```
1 1744: contig of 1744 bp in length
1745 1844: gap of 100 bp
1845 2954: contig of 1110 bp in length
2955 3054: gap of 100 bp
3055 4839: contig of 1785 bp in length
4840 4939: gap of 100 bp
4940 7461: contig of 2522 bp in length
7462 7561: gap of 100 bp
7562 11912: contig of 4351 bp in length
11913 12012: gap of 100 bp
12013 15127: contig of 3115 bp in length
15128 15227: gap of 100 bp
15228 19176: contig of 3949 bp in length
19177 19276: gap of 100 bp
19277 24838: contig of 5562 bp in length
24839 24938: gap of 100 bp
24939 30632: contig of 5694 bp in length
30633 30732: gap of 100 bp
30733 37821: contig of 7089 bp in length
37822 37921: gap of 100 bp
37922 44638: contig of 6717 bp in length
44639 44738: gap of 100 bp
44739 49424: contig of 4686 bp in length
49425 49524: gap of 100 bp
49525 56517: contig of 6993 bp in length
56518 56617: gap of 100 bp
56618 65413: contig of 8796 bp in length
65414 65513: gap of 100 bp
65514 75659: contig of 10146 bp in length
```

\* 75660 75759: gap of 100 bp  
 \* 75760 86433: contig of 10674 bp in length  
 \* 86434 86533: gap of 100 bp  
 \* 86534 98763: contig of 12230 bp in length  
 \* 98764 98863: gap of 100 bp  
 \* 98864 109919: contig of 11056 bp in length  
 \* 109920 110019: gap of 100 bp  
 \* 110020 125264: contig of 15245 bp in length  
 \* 125265 125364: gap of 100 bp  
 \* 125365 141043: contig of 15679 bp in length  
 \* 141044 141143: gap of 100 bp  
 \* 141144 170896: contig of 29753 bp in length.

## FEATURES

source  
 1. .170896  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="RP11-6P16"  
 /clone\_lib="RP11-6P16"  
 /clone="RP11-6P16"  
 /note="assembly\_fragment"  
 1845. .2954  
 /note="assembly\_fragment"  
 3055. .4839  
 /note="assembly\_fragment"  
 4940. .7461  
 /note="assembly\_fragment"  
 7562. .11912  
 /note="assembly\_fragment"  
 12013. .15127  
 /note="assembly\_fragment"  
 15228. .19176  
 /note="assembly\_fragment"  
 19277. .24838  
 /note="assembly\_fragment"  
 24939. .30632  
 /note="assembly\_fragment"  
 30733. .37821  
 /note="assembly\_fragment"  
 37922. .44638  
 /note="assembly\_fragment"  
 44739. .49424  
 /note="assembly\_fragment"  
 49525. .56517  
 /note="assembly\_fragment"  
 56618. .65413  
 /note="assembly\_fragment"  
 65514. .75659  
 /note="assembly\_fragment"  
 75760. .86433  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:left  
 86534. .98763  
 /note="assembly\_fragment"  
 98864. .109919  
 /note="assembly\_fragment"  
 110020. .125264  
 /note="assembly\_fragment"  
 125365. .141043  
 /note="assembly\_fragment"  
 141144. .170896  
 /note="assembly\_fragment"  
 41547 a 41714 c 41241 g 44377 t 2017 others

## ORIGIN

Query Match 74.4%; Score 18.6; DB 61; Length 170896;  
 Best Local Similarity 84.0%; Pred. No. 3.9e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggatctgctgtttgaagcgcagcag 25  
 |||||

Db 95897 GGATCTGCTGTTTCATGAGCAGCAG 95873

## RESULT 44

LOCUS AP000957/c  
 DEFINITION Homo sapiens genomic DNA, chromosome 21q21.1-q21.2 clone:B781M3, L156-APP region, complete sequence.

ACCESSION AP000957

VERSION AP000957.2 GI:7077200

KEYWORDS HTG.

SOURCE Homo sapiens DNA, clone:B781M3.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 190937)  
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

## TITLE

Published Only in Database (1999) In press

## REFERENCE

2 (bases 1 to 190937)  
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,  
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

## TITLE

Submitted (17-DEC-1999) to the DDBJ/EMBL/GenBank databases.  
 Masahira Hattori, The Institute of Physical and Chemical Research  
 (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1  
 Kitasato, Sagamihara, Kanagawa 228-8555, Japan  
 (E-mail:hattori@sc.riken.go.jp, URL:http://hdp.gsc.riken.go.jp/,  
 Tel:81-42-778-9923, Fax:81-42-778-9924)

## COMMENT

On Feb 24, 2000 this sequence version replaced gi:6997410.

## FEATURES

Location/Qualifiers

1. .190937  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="21"  
 /clone="B781M3"  
 /map="21q21.1-q21.2"

BASE COUNT 62556 a 34315 c 34595 g 59471 t

## ORIGIN

Query Match 74.4%; Score 18.6; DB 91; Length 190937;  
 Best Local Similarity 84.0%; Pred. No. 3.9e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ggatctgctgtttgaagcgcagcag 25  
 |||||

Db 83614 GGATCTGCTGTTTGAAGAGCTGCAG 83590

## RESULT 45

LOCUS AC016821  
 DEFINITION Homo sapiens chromosome 10 clone RP11-404C6, WORKING DRAFT  
 SEQUENCE, 16 unordered pieces.

ACCESSION AC016821

VERSION AC016821.6 GI:13605968

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

## SOURCE

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 201652)  
 Smith,D.R.

## TITLE

Genome Therapeutics Corporation Sequencing Center: Human Genome

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 201652)  
 Smith,D.R.

## TITLE

Direct Submission

## JOURNAL

Submitted (08-DEC-1999) Genome Therapeutics Corporation, 100 Beaver  
 Street, Waltham, MA 02453, USA

COMMENT On Apr 12, 2001 this sequence version replaced gi:9887636.

```

----- Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
----- Project Information
Center project name: hg066
----- Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 189359 bases at least Q40
Consensus quality: 192821 bases at least Q30
Consensus quality: 194289 bases at least Q20
Insert size: 200251; sum-of-contigs
Quality coverage: 6.6x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1116: contig of 1116 bp in length
* 1117 1216: gap of unknown length
* 1217 2340: contig of 1024 bp in length
* 2241 2340: gap of unknown length
* 2341 3504: contig of 1164 bp in length
* 3505 3604: gap of unknown length
* 3605 4932: contig of 1328 bp in length
* 4933 5032: gap of unknown length
* 5033 6078: contig of 1046 bp in length
* 6079 6178: gap of unknown length
* 6179 7870: contig of 1692 bp in length
* 7871 7970: gap of unknown length
* 7971 9374: contig of 1404 bp in length
* 9375 9474: gap of unknown length
* 9475 12523: contig of 3049 bp in length
* 12524 12623: gap of unknown length
* 12624 19429: contig of 6806 bp in length
* 19430 19529: gap of unknown length
* 19530 26456: contig of 6927 bp in length
* 26457 28556: gap of unknown length
* 28557 35832: contig of 9276 bp in length
* 35833 35932: gap of unknown length
* 35933 51442: contig of 15510 bp in length
* 51443 51542: gap of unknown length
* 51543 76692: contig of 25150 bp in length
* 76693 76792: gap of unknown length
* 76793 116535: contig of 39743 bp in length
* 116536 116635: gap of unknown length
* 116636 159761: contig of 43126 bp in length
* 159762 159861: gap of unknown length
* 159862 201652: contig of 41791 bp in length.
*
* Location/Qualifiers
* 1..201652
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="10"
* /clone="RP11-404C6"
* /clone_lib="RPCI-11"
* 1..1116
* /note="assembly_name:Contig2"
* 1217..2240
* /note="assembly_name:Contig3"
* 2341..3504
* /note="assembly_name:Contig7"
* 3605..4932
* /note="assembly_name:Contig11"
* 5033..6078
* /note="assembly_name:Contig15"

```

```

misc_feature 6179..7870
              /note="assembly_name:Contig25"
misc_feature 7971..9374
              /note="assembly_name:Contig26"
              clone_end:SP6"
misc_feature 9475..12523
              /note="assembly_name:Contig27"
misc_feature 12624..19429
              /note="assembly_name:Contig28"
misc_feature 19530..26456
              /note="assembly_name:Contig29"
misc_feature 26557..35832
              /note="assembly_name:Contig30"
              clone_end:T7"
misc_feature 35933..51442
              /note="assembly_name:Contig31"
misc_feature 51543..76692
              /note="assembly_name:Contig32"
misc_feature 76793..116535
              /note="assembly_name:Contig33"
misc_feature 116636..159761
              /note="assembly_name:Contig34"
misc_feature 159862..201652
              /note="assembly_name:Contig35"

BASE COUNT 50230 a 47697 c 49286 g 52927 t 1512 others
ORIGIN
Query Match 74.4%; Score 18.6; DB 64; Length 201652;
Best Local Similarity 84.0%; Pred. No. 3.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ggatctgctgttttgaagcgcagcag 25
   ||||| ||||| ||||| |||||
Db 147871 GGATCTGCTGTTTCATGAGCAGCAG 147895

```

Search completed: October 9, 2001, 12:11:18  
Job time: 3793 sec

FEATURES  
source

```

misc_feature 1..1116
              /note="assembly_name:Contig2"
misc_feature 1217..2240
              /note="assembly_name:Contig3"
misc_feature 2341..3504
              /note="assembly_name:Contig7"
misc_feature 3605..4932
              /note="assembly_name:Contig11"
misc_feature 5033..6078
              /note="assembly_name:Contig15"

```





